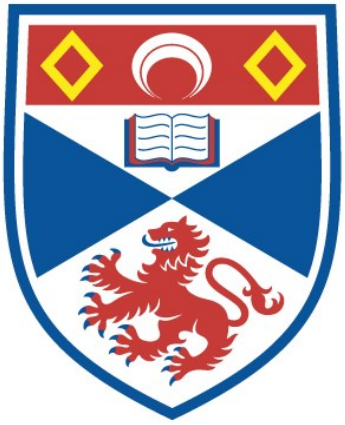
MORAL REALISM, MORAL EXPERTISE AND PATERNALISM

Gudmundur Heidar Frímannsson

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

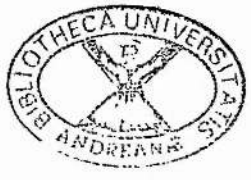


1992

MORAL REALISM, MORAL EXPERTISE AND PATERNALISM

Gudmundur Heidar Frímannsson

PhD thesis
submitted in
the Department of Moral Philosophy
University of St. Andrews
June 1991



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

In this essay I examine the notion of moral objectivity of moral properties. Moral objectivity seems to be able to resist the arguments of subjectivists. There seem to be true moral sentences and moral facts can explain actions and occurrences in the world. Values seem best accounted for in objective terms and persons can have interests or good independently of their desires. It seems to be reasonable to think of the nature of moral value in terms of consequences. Knowledge requires truth so the objectivity of moral properties makes moral knowledge possible. Moral knowledge should be accounted for in similar terms as other kinds of knowledge. The major requirement on moral knowledge is coherence. Moral expertise is both possible and plausible and so are moral experts. Paternalism is possible because our values can conflict: autonomy can conflict with general welfare. Paternalism is making someone do what is in his own interest. This seems best thought of in terms of the consequences for his good. The justification of paternalistic interventions seems best based on the weighing of the consequences of the intervention and the decision of the agent. One thing which must be taken into this weighing is the rationality of the decision of the agent. Rationality is basically thought of as the maximization of good. Autonomy is part of everyone's good. It can conflict with the agent's general or overall welfare. But the importance of autonomy for every agent creates a presumption against paternalism. But paternalism can maximize autonomy and paternalism can be justified to secure some minimal autonomy. So paternalism and autonomy seem to be compatible.

Acknowledgements

Various people have assisted me when working on this thesis. My three supervisors were very helpful. First it was dr. Gordon Graham who suggested in the earliest stages when I was contemplating objectivity that I tried to make the link between paternalism and objectivity central in the thesis. The second, mr. Anthony Ellis, commented on the draft which turned into the first chapter and the first two sections of the second. Thirdly I have benefitted greatly from professor John Skorupski's incisive comments on the whole thesis and his generous encouragement. While working on the thesis I received a an ORS award from the Committee of Vice-Chancellors and Principals of the Universities of the United Kingdom and a research grant from the University of St. Andrews for which I am grateful. I also wish to express my gratitude to my family, to my two children, Jóhann and Valgerður, and my wife, Elísabet, for their patience and consideration while I was working on this thesis.

Declarations

I, Gudmundur Heidar Frímannsson, hereby certify that this thesis, which is approximately 85,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

10 June 1991 Gudmundur Heidar Frímannsson

I was admitted as a research student under Ordinance No. 12 in November 1987 and as a candidate for the degree of PhD in December 1987; the higher study for which this is a record was carried out in the University of St. Andrews between 1987 and 1991.

10 June 1991 Gudmundur Heidar Frímannsson

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

10 June 1991 Professor John Skorupski

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

10 June 1991 Gudmundur Heidar Frímannsson

Contents

Introduction

The basic idea in this essay is, hopefully, simple. It seems to be true that we can make mistakes about our own interests and about moral rightness and wrongness. We are sometimes plain wrong about our best interests and we sometimes neglect to help people in need. It may also seem plausible to explore the idea that it may be right, at least sometimes, to make an agent do what is in his own interest if he was about to do something harmful. In this essay I attempt to work on the idea of objectivity of moral values and the truth conditions of moral propositions to explain what is involved in this idea of something being in our interest or something being morally right independently of our awareness and knowledge.

The connection between making someone do what is in his own interest or preventing him from committing a major mistake and objectivity is not quite straightforward. It seems that in addition to objectivity we need a substantive notion of moral values. It also seems to require that we can say that there is knowledge of what is morally right and good and that there are some agents who are in possession of this knowledge. With this on board it seems to be possible to make out a case for having an agent sometimes do what is in his own best

interest and at the same time accepting that in general he is the best judge of his own interest.

This view of objectivity is certainly not unchallenged. Subjectivists of any sort have tried to challenge this view of moral properties. They are not willing to accept that moral properties are objective or that moral propositions can have truth value. This seems to have the consequence that no moral mistakes are possible. This would be an unacceptable consequence. But subjectivism can be a subtle doctrine and subjectivists have wanted to avoid this consequence.

In the first chapter I consider two subjectivist theories. The first is R. M. Hare's distinction between descriptive and evaluative meaning. This distinction is fundamental to Hare's whole theory. If it can be rejected, then many other claims of his theory seem to follow. It seems to me that this distinction is not tenable. It does not seem to be possible to weed out the evaluative meaning from the other parts of the meaning of terms like 'good' or 'murder'.

Simon Blackburn has developed a robust theory of morality assuming subjectivism but at the same time taking moral language at face value, i.e. accepting that moral propositions can have truth value and can play a proper role in inferences in conditionals. This account seems to be objectionable on basically two grounds. First, it assumes a strong distinction between first- and second-order questions, between ethics and metaethics. This distinction is useful but it does not seem to hold

in the strong version required by Blackburn. Second, it seems that some elements of his theory cannot succeed on their own terms.

In the second chapter I look at four theories which can broadly be called moral realist. It seems to me that they all must be rejected. It is not that everything is wrong in them but they seem to have various unsatisfactory features. The view of moral qualities I come to is broadly realist and that moral propositions can have truth value. I also argue for two more conclusions in this chapter. The first is that we should accept consequentialism as a substantive account of the nature moral value. The second is that moral values are objective.

Knowledge requires truth. Moral propositions have truth values and it seems reasonable to accept that some of them are true. Hence, moral knowledge seems to be possible. On the best possible justification of a claim to knowledge there is always the possibility that the claim might be false. So justification does not guarantee knowledge. Neither seem there to be any good candidates for self-evidently true foundations for knowledge. So fallibilism seems inevitable and a coherence theory of justification for knowledge. The same applies to moral knowledge. The way to acquire moral knowledge is through critical reflection.

The fourth chapter explores the possibility of moral expertise and moral experts. The notion of moral expertise has not been thoroughly researched by

philosophers but it is strongly resisted. The main reason for the resistance, it seems, is that morality requires the rational agent to be responsible for his own actions. It has been widely believed that moral expertise undermines responsibility. This seems false. A rational moral agent requires knowledge for his actions. Under some circumstances he might require moral expertise from others. Moral experts are competent judges in morality and they should be able to formulate opinions and theories about morality, especially the part of morality which can be usefully considered new.

Paternalism is the subject of the fifth chapter. Paternalism is to be made to do something in one's own best interest. There are varieties of ways of making agents do or accept what is in their own interest. Some ways are not paternalistic. Rational persuasion without any untoward pressure is not paternalistic. But coercing someone to wear seatbelts or to take physical exercise is paternalism if the reason for the coercion is the agent's own interest. Whether an intervention is paternalistic or not is independent of its justification. In the justification of a paternalistic act it is necessary to weigh the consequences of allowing the agent to decide against the consequences of interfering. The rationality of the agent's decision must also be taken into account. The aim of the interference is to maximize the agent's good.

In the last chapter I discuss autonomy. Autonomy is the agent's ability to see and act on good reasons. This

requires knowledge of everything concerning the agent's own interest and the agent must have control over his life. It seems reasonable to believe that autonomy could exclude paternalism altogether. One way of arguing for this might be in terms of the notion of side constraints. Side constraints may be taken to exclude all undue interference. But it does not seem to be the right conclusion to exclude interference come what may. Autonomy puts a limit on paternalism. Paternalism can only be accepted as a prevention against great harm and when the agent does not reach a certain minimum of rationality. Paternalism seems also to be justified to maximize autonomy.

One thing needs to be cleared up at the start. The notion of objectivity is obviously central to the whole essay. It appears in three contexts which might be thought to be contradictory. In the first instance moral properties are said to be objective in the sense of corresponding to true moral propositions. This is most prominent in the first two chapters. In this sense moral properties are independent of minds. The second context is when I discuss desires and value in 2.3. I argue for an objectivist conception of value as against a desire-satisfaction theory of value. The basic idea is that our values cannot be explained by the satisfaction of our desires. This idea is analogous to the notion of moral properties argued for earlier. But in 6.4. I say that autonomy is valuable not as a Platonic Form, discoverable in a reality independent of human desires and

consciousness, but it is valuable because if rational agents were subjected to the experiences of autonomy they would desire it. This seems to clash with what had been said earlier about moral properties and values.

Is there a contradiction here? To answer that question we must first say what the contradiction in question is supposed to be. The contradiction is this: in discussing moral properties they are said to be independent of human minds or independent of being recognized. When talking about autonomy as a part of human welfare I seem to accept that its value can be interpreted as dependent on counterfactual or informed desire.

This contradiction depends on how we understand the notion of a moral property. How is it related to moral values? One answer to this question might be: moral value is a subset of moral properties and nothing is implied about moral properties in general by saying that some part of moral value should be explained by counterfactual desires. Another answer might be this: moral value in this essay is used in a loose sense to cover both typical values like goodness and requirements like rightness. It coincides roughly with the notion of a moral property. Explaining moral value is explaining moral properties. It is an argument for this interpretation that right actions are considered to be a subset of good actions. This second interpretation seems to be the correct one. But as should be clear by now, this interpretation seems to have

the immediate consequence that what I said about autonomy and moral properties is contradictory.

Yet, when looked at more closely there is no contradiction. The reason is this. When I argue that autonomy is such that it would be desired were it experienced by rational agents I do not mean to imply that autonomy is valuable because it would be desired if this condition were fulfilled. It is because rational agents are as they are that they would desire autonomy were they to experience it. And I think it is a fact that they do desire it. But if, for some reason, rational agents were to change and stop desiring autonomy if they experienced it then autonomy would not stop being a part of objective welfare. This welfare seems to be a necessary part of a good life. It is fortunate that rational agents are so constituted that they do desire autonomy if they get the chance to experience it. We can say that rational agents ought to desire autonomy as a part of their welfare but there can be circumstances where this ought is defeasible. Autonomy is not the supreme value. I do not have a theory of a unified order of values. So even though rational agents changed autonomy would continue to be a part of their objective welfare. It is therefore true of autonomy as it is true of moral values in general that they can transcend our recognition of them. So there is no contradiction between my views on the status of autonomy and the nature of moral properties.

These are the main lines of argument in this essay. Moral mistakes are possible. It is not only individuals who make them. Whole communities can systematically make moral mistakes. If a community changes and starts taking the torturing of children as acceptable behaviour, then it has become corrupted. This intuition is respected by what I argue in this essay. If there were moral experts in such a society and they accepted this conduct they would be wrong. There is no guaranteeing the truth of our considered beliefs.

Chapter 1

Objectivity

In this first chapter I start by drawing some important distinctions for discussing moral realism and give a rough description of it. Then I go on to consider two challenges to moral realism. The first is Hare's theory of prescriptive and descriptive meaning. The second is Simon Blackburn's quasi-realism. I try to argue that we have reason to reject these two theories.

1.1. Some Distinctions

Goodness is a value. It seems that there can be three kinds of goodness. First, there can be goodness of a kind. This means that if we talk about a good knife we mean that the knife is a good example of a knife. What is a good example of a knife is determined by standards and the purpose we are using it for. This notion could be analyzed further and we could certainly find other kinds of goodness similar to this one.[1] Second, there is prudential goodness. This is goodness of things which can satisfy our self-regarding needs or fulfill our self-regarding desires. Yet, not all things which can fulfill those needs and desires are prudentially good. Only those

1. G.H. von Wright: The Varieties of Goodness, London, Routledge and Kegan Paul, 1963, see esp. the first 5 chapters, pp.1-113

things which can be said to be in our own interest can be counted as prudentially good. All decisions based on prudence are aimed at enhancing or promoting our own interests. Thirdly, there is moral goodness. What is moral is determined by what can be sanctioned. It is an instance of moral goodness when we promote the interest of others. If someone is in distress and we can alleviate his distress without undue effort, we ought to do it. If we do not do it we feel guilty and guilt is a sanction.

Sometimes 'value' is used to cover various different things like aesthetic value, duty etc. beside goodness.[2] If used in this sense requirements like rightness become a type of value. But usually a distinction between goodness and rightness is drawn in terms of deontology and axiology: the former deals with requirements like rightness, the latter with values like goodness. However we distinguish between these two it is important to realize how goodness and rightness are related. We can describe a moral agent as an agent that is concerned with the right kind of thing for the right kind of reason. Now how should the meaning of 'right' be explained? One suggestion might be that the right thing to do is what is our duty to do. But 'duty' has no clearer meaning than 'right'. So it would be reasonable to expect an explanation of the meaning of 'duty'. How should we explain its meaning? I suggest that its meaning should be explained as the best action out of the set of actions

2. See Frankena, W.K.: "Value and Valuation" in P. Edwards (ed.): The Encyclopedia of Philosophy, New York, Macmillan Publisher, 1963, vol. 8, p. 229

available to the agent at any particular time. This has the obvious consequence that right actions become a subset of good actions. Now this may not be thought to be very illuminating and it is possible to point to differences between these two concepts which emerge, for example, in contract theory of justice.[3] But this difference is not important for this essay.

The bearers of moral goodness are agents and actions. If we talk about a good man, this is usually taken to involve a virtuous man. It does not exclude that he does no evil or bad in his whole life, but only that there are reasons explaining it when he does and on the whole he does more good than bad.

Actions can be done for prudential reasons or moral ones; this distinction is not exhaustive. It seems to follow from the distinction drawn earlier that no moral actions can only or primarily concern the agent. The distinction is that prudential reasons primarily concern the agent himself but moral reasons concern other people. The obvious first suggestion for moral goodness of an action is when it benefits another person and moral badness when it harms him.[4] But there are two problems with this suggestion as it stands. Moral goodness is not relative to persons. It is either good or bad simpliciter, not good for me or bad for me or good or bad for somebody. But the relations of harming and benefitting are obviously relative to persons. The other

3. J. Rawls: A Theory of Justice, Oxford, Oxford University Press, 1971, pp.446-452
4. G.H. von Wright: op. cit. p. 119-130

problem is that my action can benefit another person without it being an instance of moral goodness. The same applies to moral badness. Let us assume that I encourage somebody to drink liquid in a glass to cure his coughing, believing it to contain cyanide and that it will really kill him. But to my surprise it turns out that it contains lemon tea which does wonders for his coughing. My action is not an instance of a morally good action, even though it benefitted him. It is actually an instance of a morally evil action because I intended to kill him. The notion of intention is obviously crucial in this field. So some account of intention is needed to give a full description of a good action. But it will have to suffice to say that the basic sense of 'good intention' is doing good.

This is where the notion of sanctions comes in. Only those actions that can be sanctioned can be moral. Sanctions can be either internal, like guilt, or external, like being put into prison or social disapproval. This, obviously, puts no limit on what type of actions can be moral, given that they can be sanctioned. This applies both to harming and benefitting. If we omit to benefit someone out of negligence then we ought to feel guilty. If we harm someone intentionally, then we ought to feel guilty and we ought to be punished. This possibility of sanction explains why we found problems with the suggestion above that moral goodness consisted in benefitting someone and why the notion of intention plays a central role in a moral action. Another

consequence of making sanctions central to morality is that the other-regardingness of moral actions is not a necessary feature of them, as may have been thought from what I said above. It is still true that other-regarding actions are a primary example of moral actions, but sanctions do not exclude the possibility of self-regarding actions as being moral. This has, at least, the virtue of historical ancestry because many moral theorists have considered actions regarding oneself as morally good or bad. Suicide has been thought to be morally blameworthy.

Using the notion of sanction in this way goes along well with many of the traditional Christian moral rules which are prohibitions on doing wrong: do not commit adultery, do not kill, do not covet. Other instances of moral conduct fit in with this as well: "wreaking vengeance; responding hubristically to being elevated to power; responding cantankerously to authority; responding to another's achievement by detraction, slander or other behaviour expressive of envy."[5] Refraining from this type of conduct is clearly an instance of moral goodness. Indulging in all these types of conduct should invite guilt. So sanctions seem to be necessary for the moral.

It might be objected to this that it is not the possibility of sanctions that is necessary for the moral but the possibility of being rightly sanctioned. And if this be so, then I seem to be assuming what I am trying

---

5. Holland, R.F.: "Good and Evil in Action" in Against Empiricism, Oxford, Basil Blackwell, 1980, p. 112

to explain. In response it is fair to point out that I did not claim that the possibility of sanctions was sufficient. The reason was that I am sceptical about the possibility of defining morality. It seems to me to be true that something along similar lines to those suggested in the objection would have to come in a full definition. But I shall not examine this possibility any further.

Another objection to using the notion of sanction in this way could be this: if sanctions are necessary for morality then moral realism is thrown in doubt. The reason is that the moral does not seem to be independent of the mental if sanctions are necessary. As I indicate below it seems to be one feature of moral realism that moral facts are believed to be independent of mental states. How should this be handled? There is an analogy between the moral and the semantic in this respect at least: the community of speakers seems to be a necessary condition for the semantic but this in no way implies that the semantic depends on the existence of a community of speakers. The same thing could be said about sanctions and the moral. Sanctions make the moral possible but the moral does not depend on sanctions but is independent of them.

Moral requirements seem to be such that they apply to all agents in relevantly identical situations. This can be taken to mean two things. First, it can mean that moral rules are universal in character, there are no references to particular individuals. Second, it can mean

that moral rules are agent neutral; "if it is a reason for anyone to do or want something that it would reduce the amount of wretchedness in the world, then that is a neutral reason."[6] The difference between these two can be illustrated by the principle of rational egoism: Everyone has a reason to promote his own interest. This principle is universal, but it is not agent neutral because it involves reference to the interest of every agent, it is agent-relative. An agent-neutral principle would say that if x is a part of A's good, then everyone has a reason to promote it.[7] It seems to me that the moral includes generality in both these senses.

Moral realism is a particular view about the nature of moral and non-moral goodness. It is not quite obvious how it should be formulated. One natural suggestion is that the moral is objective. But this is really no clarification unless we have some account of objectivity. One idea for such an account might be that moral statements have truth values, can either be true or false. But there are at least two reasons for not accepting this as sufficient. First, all moral statements might be false. This would invalidate moral realism. We must have a reason for thinking that some moral statements are true, if we are to accept this suggestion about moral realism. Second, the truth of some moral statements might be dependent on some facts about the

---

6. Nagel, T.: The View from Nowhere, New York, Oxford University Press, 1986, pp. 152-153
7. Skorupski, J.: John Stuart Mill, London, Routledge, 1989, p. 309

mental. We could hold, for example, that relieving wretchedness in the world is a moral reason for everyone to act because every rational agent actually has a desire to relieve wretchedness in the world. So this is a moral truth because everyone has this desire. This is certainly a possible version of moral realism. But I think it is true to say that moral realism is often taken to imply that the moral is independent of the mental. So I do not think this is a satisfactory account of moral realism.

Let us try another formulation of moral realism. Moral realism must include two claims: One is that there are moral truths in the sense that there are true moral propositions, and two, these truths are independent of our beliefs or evidence for them. These two claims are certainly important parts of any moral realism but they are not sufficient to exclude some views that have not been generally accepted as moral realism. Moral realism of this sort excludes moral constructivism which accepts moral truth but claims that it is constituted by the evidence for the truth and it excludes moral noncognitivism which does not accept that moral statements can have truth value. But moral realism, as described here, does not exclude a theory like rational egoism and would have to accept it as morally realistic. The reason is that propositions about the psychological can have truth value. Therefore, moral values might be subjective in an important sense according to this moral realism.

It follows from this that there is a difference between objectivism about values and moral realism. Moral realism is compatible with some subjectivity about values. In so far as this value-subjectivity does not involve the claim that the value of x is constituted by the belief that x is valuable, it is compatible with moral realism. It would take some time to work out where the line between an acceptable form of value-subjectivity and an unacceptable form lies but I do not intend to attempt to find it. But in 2.3. I will try to argue for an objective view of moral values.

There seem to be two major considerations that motivate arguments for moral realism. The first is that moral realism can accept moral statements as literally true if true. Moral terms enter statements of all sorts and among them are indicative statements. Moral realism accepts such statements and thinks that they refer to moral properties. It does not need to resort to some subtle explanations of moral terms as not really referring to anything or having some other kind of meaning or explaining them away in some other fashion. I will have a look at two such attempts later in this chapter. But this fact about moral realism that it accepts ordinary language at face value, so to speak, gives it the initial advantage of putting the onus of proof on other theories. The other major consideration for moral realism is that it accords well with realism about other kinds of discourse like scientific discourse. The force of this argument obviously depends on how

plausible one finds realism generally. I do not want to enter into arguments about scientific realism, but I only want to indicate that many of those who have been willing to accept realism about science have been very critical of such a view about moral discourse. Yet if one accepts scientific realism, then moral realism becomes more plausible and if one thinks that moral realism cannot be accepted but scientific realism can, then one needs specific arguments for this view on moral realism.

Scientific realism accepts that scientific propositions are true independently of our evidence for them or our perceptions of the relevant states of affairs. Both moral and scientific realism are objective in this sense. This indication, and the reasoning above, about the connection between objectivism and moral realism seem to me to be sensible. But there is another question about moral properties that should be discussed at this stage. It is whether moral properties should be understood naturally, nonnaturally or supernaturally. According to the last the moral property of being good would consist in being willed by God or some other supernatural being. The merits of this conception I will not discuss, not because it is without any merit but because it is more plausible to think that God wills what is good rather than that what is good is good because it is willed by Him. Discussion of this issue is outside the scope of this essay. This leaves us with the other two possibilities.

Many modern moral philosophers have argued that a moral property could not be a natural property. The argument for thinking that was basically simple. If moral properties were natural properties the relation of synonymy between 'good' and some name of a natural property ought to obtain. But the problem is that no such term seems to be synonymous with 'good'. We can always ask about any natural property: is it good? The conclusion is that goodness is a sui generis property or a nonnatural property. But this view carries with it problems. One is how we are supposed to know nonnatural moral properties. Proposing a special moral faculty enabling rational agents to know what is morally good and right is implausible. This seems to make moral realism a very difficult or nearly an impossible position.

One way out of this is to realize that moral properties can be natural properties without the relation of synonymy obtaining between the terms referring to them. Notwithstanding this possibility however, one may prefer to argue that moral properties can be constituted by natural properties. This does not mean that moral properties are identical with natural properties. The relation between them will be the relation of supervenience. To hold that moral properties supervene on natural properties is to say that if the natural properties of two things are identical, then the moral properties cannot differ. There is much more to be said about this conception of natural properties and I will try to say some of those things in this and the next

chapter. But, given that what I have to say against Blackburn's theory of quasi-realism is reasonably successful, I think it is true to say that any of these three construals of moral properties, as natural, nonnatural or supernatural, would be realistic.

I shall attempt to say something more about the form that moral realism should take in the next chapter. But before I do that I want to consider two important challenges to moral realism. In trying to answer some of the criticisms in these two views, I hope to clarify important features of moral realism.

1.2. Hare's Challenge

Hare argues for noncognitivism in the sense that he denies that evaluative terms or propositions including evaluative terms can have truth value in virtue of the evaluative element in them. He is not a straightforward noncognitivist because he believes that value terms can have descriptive meaning. In so far as they have descriptive meaning they can have truth values. But the crucial point is that nothing can be inferred about evaluation from description. I want to have a look at his argument for this theory.

We can distinguish between three things concerning sentences. The first is **force**. A sentence can have assertoric force, imperative force and so on. But the same sentence can be used as an assertion, an imperative, a question. So force cannot be a property of sentences.

It is a property of speech acts. An utterance of a sentence for a purpose is a speech act. Examples are assertions, questions and commands. The second is **mood**. The mood of a sentence can be indicative, imperative or subjunctive. Sentences in the indicative mood (or to be exact containing verbs in the indicative mood) usually state facts, in the imperative they usually express orders. An indicative uttered by an actor on stage is not taken to be an assertion of anything. Yet it is an indicative. So mood is a property of sentences. The third thing is **sense**. The sense of a sentence is the way of determining its reference. The importance of making these distinctions is to realize that only mood and sense can be a part of what is usually called the meaning of a sentence.

It is natural say about descriptive sentences that they are indicatives. Indicative sentences which are not descriptive would seem to be a very peculiar type of sentences. We assume that only properly descriptive sentences can have truth value.[8] The argument for taking some indicative sentences as not being descriptive can be that otherwise we would be confusing logical form with grammatical form. When we considered evaluative terms we would come to the conclusion that it is seriously misleading to think that sentences including evaluative

---

8. McGinn, C.: "Semantics for Non-Indicatives" in Philosophical Studies 32, 1977, p. 303 where he says: "...and if the mood of the sentence is not indicative it still cannot, without impropriety, be said to possess truth-value"

terms should be categorized along with other indicative sentences.

Let us spell out the argument for this in some detail. The first thing to realize is that the question is not about the terms themselves but their meaning according to Hare. Value terms have two kinds of meaning: descriptive and evaluative. And there is a sharp distinction to be drawn between these two. You cannot draw any evaluative conclusion based on descriptive premisses. This argument implies that all indicative sentences have descriptive meaning but some have evaluative meaning as well. The obvious question is: How do we know which ones have evaluative meaning? The answer cannot be: we know which ones have evaluative terms because we cannot characterize those terms without recourse to their meaning. The answer to this question implied by Hare's argument is that we do things with sentences having evaluative meaning which we cannot do with descriptive meaning. This needs some explaining.

It seems to be unproblematic to distinguish between describing and evaluating. These are two kinds of activities we engage in with language and someone who knows how to speak must know how to engage in these two kinds of activities. So knowing the language entails being able to distinguish between describing and evaluating. But this is not the same as accepting the sharp distinction here described between descriptive

meaning and evaluative meaning. To see that we need to go through the argument for the distinction.[3]

When we say that a certain wine is good, Hare calls it Colombey-les-deux-églises 1972, the reason we give for thinking a wine good is that it has a particular taste, bouquet, body, strength etc. His idea is to try to isolate the evaluative meaning of a term from the descriptive meaning and show that the distinction is valid.

> "Let us invent a word, 'ϕ', to stand for that quality of the wine which makes us call it a good wine. The quality is, as I have explained, a complex one. Will you allow me to suppose, also, that (as is not improbable) by the time 1972 wines of this sort begin to be good, the science of aromatics (if that is the right name) will have advanced enough to put the wine snobs out of business; that is to say, that it will have become possible to manufacture by chemical means additives which, put into cheap wines, will give to them tastes indistinguishable by any human palate from those of expensive wines. We should then have a chemical recipe for producing liquid tasting ϕ. This would make it easy (though even without such scientific advance it would be

3. Hare, R.M.: "Descriptivism", in Hudson ed.: The Is-Ought Question, London, Macmillan, 1969, pp. 241-247. Further arguments for Hare's position are to be found in R.M. Hare: Freedom and Reason, Oxford, Clarendon Press, 1963, p. 26-29, 152-156; R.M. Hare: The Language of Morals, Oxford, Clarendon Press, 1952, p. 111, 118-125, 146-150.

perfectly possible) to teach somebody to recognize the ϕ taste by lining up samples of liquids tasting ϕ, and others having different tastes, and getting him to taste them, telling him in each case whether the sample tasted ϕ or not. It is worth noticing that I could do this whether or not he was himself disposed to think that these liquids tasted good, or that, if they were wines, they were good wines. He could, that is to say, learn the meaning of 'ϕ' quite independently of his own estimation of the merit of wines having that taste.

...'I want you to understand that, in calling a wine ϕ, I am not thereby commending it or praising it in any way, any more than it is commending it or praising it to say that it is produced by this chemical recipe; I am indeed (for such is my preference) disposed to commend wines which have this taste; but in simply saying that wine is ϕ I am not thereby commending it any more than I should be if I said that it tasted like vinegar or like water. If my preference (and for that matter everybody else's) changed in such a way that a wine tasting like this was no longer thought any good, and we could do nothing with it but pour it down the drain, we could still go on describing it as ϕ.'"[10]

Hare claims that someone who does not accept the distinction between descriptive and evaluative meaning

10. Hare, R.M.: op. cit. p. 242-243

cannot understand the argument above. The example he chose, he claims, was perfectly natural and valid and we could not do without the distinction in trying to understand it. Hence it makes no sense to refuse to accept it. As the example showed it is perfectly possible for two people to agree about whether Colombey 1972 tastes ϕ and disagree about whether it is a good wine. This would not be possible on the premise that there was no distinction between descriptive and evaluative meaning. He rejects the charge that he has assumed what he wanted to prove by introducing a distinction between commendation and description in the argument. He had only put into it the premise that the distinction was intelligible. And because it was intelligible and the argument went through it was valid.

This is an important argument, if true. I think it can be seriously doubted. To see why we should take a careful look at the assumptions. There seem to be three key assumptions here:

1. To call x good is to commend it.
2. Being good is directly related to our preferences.
3. There is an unbridgeable gap between the sentence "x is good" and "x has a certain taste, bouquet, body, strength etc."

Let us first look at the third assumption. What Hare does in effect in the example and the argument is to coin the name ϕ for the latter sentence in the third

assumption and say it does not mean the same as the former sentence, i.e. "x is good" does not mean the same as "x has certain taste, bouquet, body, strength etc.". It is perfectly true that there is no analytic connection between these two sentences. Yet there is a connection. The qualities described in the conjunctive sentence are criteria for the goodness of wine. If all of the qualities mentioned were absent, the wine couldn't possibly be good. But if they are all present, it certainly is. This shows that it is the presence of those qualities which makes the wine good. The goodness of the wine consists in the presence of these qualities. So there is a relation between these two sentences but it is not analytic.

This is the real is-ought gap that is sometimes believed to be fatal to moral realism. The gap between is-ought propositions is similar to gaps between is-is propositions. There is no analytic relation between biological propositions and propositions in chemistry. This does not exclude the possibility of propositions in chemistry being evidence for the truth of a proposition in biology. So even though we accepted that there were an is-ought gap, it should not be taken to exclude the possibility of an evidential relation between propositions of these two types. So moral realism could accept the is-ought gap. But as has been described, we can also think that the goodness consists in the presence of these other properties relevant to the goodness of wines. It is at least clear that the synthetic-analytic

distinction does not provide noncognitivism with any good arguments for denying such relations.

But the goodness of the wine can also refer to the pleasure to our palate a good wine gives. This goodness is obviously directly related to our preferences. But in so far as there is something for a good wine to be, some conditions to fulfill, then this goodness is not necessarily related to our preferences. It seems to me that wine can both be pleasant and have goodness of the kind explained above. Assumption 2 seems to be valid for one kind of goodness but not for another. It seems to me that Hare thinks that all goodness is similar to pleasure in the respect that all goodness is somehow consequent upon our preferences. I think this ultimately is based on his belief that to understand good is to pursue it. This is a a belief which says that there is an internal link between goodness, preference and choice. I will not inquire into whether internalism is the right theory about motives and reasons for action or externalism.[11] Both these theories have been taken to be compatible with moral realism.[12]

I have not attempted to argue that assumption 2 is false but I have given some reason to think that it is implausible as a general account of goodness but leaving it open whether it is a good account of moral goodness. Assumption 3 seems to me to be false. The first

11. Brink, D.O.: Moral Realism and The Foundations of Ethics, Cambridge, Cambridge University Press, 1989, pp. 38-43

12. See Brink, D.O. op. cit. ch. 3, and D. McNaughton: Moral Vision, Oxford, Basil Blackwell, 1988, pp. 46-50

assumption is left. What should be said about it. The first thing to be noticed is that it seems to be literally true: To call x good is to commend it. But we should be careful about words here. We are talking about **calling** x good and **commending** x. These are both speech acts and hence concern only the force of the sentence, not its sense or mood or in other words its meaning. It is simply a fallacy that good means commending. To commend x is to do something with respect to x, i.e. to perform a speech act. Because it is a speech act its assertibility conditions are derived from the point of the act. If I assert p, I might want to induce the belief that p in another speaker, or I might want to distance myself from some other belief or whatever. A particular speech act can serve a number of purposes and a particular sentence can be the content of different types of speech acts. Usually indicative sentences are used for asserting. There are certain conventional relations between speech acts and mood. Yet it does not follow that the meaning of the sentence can be analyzed in terms of the speech act. What is the relation between the meaning of a sentence and a speech act? The relation is only conventional: indicatives are ordinarily used to assert. The meaning of a sentence or parts of a sentence cannot be explained in terms of the speech act.

The upshot of this is that this argument at least does not prove that there is a distinction between descriptive and evaluative meaning. All the three

assumptions are either implausible or can be accommodated within moral realism.

This does not necessarily mean that there is no valid distinction between descriptive and evaluative meaning. We need a positive argument for the idea that there is no sharp distinction to be drawn between descriptive and evaluative meaning. I shall mention two things neither of which is conclusive. The first is a comparable example. The word "valid" as used in logic is an evaluative word. It indicates what are good arguments. It can be defined in terms which are not evaluative such as coherent, when premisses are true false conclusions do not follow, the conclusions follow from the premisses. When these conditions are fulfilled, an argument is valid. It is at least arguable that none of the defining terms can be considered an evaluative term. So this can be considered an example of defining an evaluative term without recourse to other evaluative terms.

The second is that it seems to be impossible with many evaluative terms to distinguish descriptive from evaluative meaning. What is the evaluative term in this sentence: he only thinks about himself. There does not seem to be any. But it is certainly a derogatory statement about anybody. What about the sentence: She is a very inconsiderate woman? "Inconsiderate" is a good candidate for a value term. But how should we weed out the descriptive component? The idea must be that there is a conjunctive sentence similar to the one we used in defining valid or the conjunctive in the third assumption

above. It might go something like this: She is rude generally, never takes other people's feelings into consideration when deciding something and so on. But here we should notice one thing. The conjunctive statement could just as well serve as an explanation of the evaluative meaning. Not to take other people's feelings into account is, after all, a defect of character. And it seems to me to be an appropriate candidate for the evaluative meaning. But how should the factual meaning then be explained? One plausible suggestion might be thought to be the language of physical science. But I will be brief about this suggestion: it is simply false that the sentence "x is considerate" can be translated into the terms of physical theory. Hence there is reason to be doubtful about the distinction between descriptive and evaluative meaning.

It needs to be pointed out that I have only discussed the distinction between evaluative and descriptive meaning and the example from Hare was distinctly non-moral. It is possible to say that the story might be different, if we discussed moral value rather than non-moral value and the distinction between morally evaluative statements and descriptive ones. On Hare's theory moral values have distinctive characteristics: prescriptivity, overridingness and universality. I do not want to deny that moral values have distinctive characteristics and I have already made that clear. But I do not think moving on to moral values would make any difference to what has been said here. The

distinction between the two kinds of meanings gains no obvious plausibility by narrowing it down to moral values rather than taking values generally.

1.3. Quasi-Realism

Quasi-realism is a theory developed by Simon Blackburn.[13] Its basic idea is that we can accept moral language at its face value and yet explain it with a slim, austere, Humean theory. To accept moral language at its face value is to accept that evaluative statements have assertoric force and value words seem to refer in the same way as any other words and all the other indications we have taken to imply that values are objective. It also accepts that moral experience can be such that values seem to be imposed on us or strike us as having a basis outside our own feelings and emotions. A quasi-realist can hold that "it is not my sentiments that make bear-baiting wrong; it is not because we disapprove of it that mindless violence is abominable; it is preferable that the world should be a beautiful place even after all consciousness of it ceases."[14] He can also say: "if everyone comes to think of it as permissible to

---

13. See S. Blackburn: Spreading The Word, Oxford, Clarendon Press, 1984, ch. 6, pp. 181-223; "Error and the Phenomenology of Value" in T. Honderich: Morality and Objectivity, London, Routledge and Kegan Paul, 1985, pp. 1-22; "Rule Following and Moral Realism" in Holtzman and Leich ed.: Wittgenstein: To Follow a Rule, London, Routledge, Kegan and Paul, 1981, pp. 163-187 for the main ideas of Blackburn's quasi-realism.
14. Blackburn, S.: "Errors and the Phenomenology of Value" p. 10

maltreat animals, this does nothing at all to make it permissible: it just means that everybody has deteriorated."[15] Yet quasi-realism believes that these facts can be explained by an austere Humean theory of values which assumes that the mind spreads itself on the world and in fact that there are no values to be found there, only our emotional reactions to natural facts which are developed in interaction with other people and become commitments and duties and as this socialization develops we start to believe that these things are discovered in reality.

If this theory really delivers the goods, it is no mean feat. Let us see how it works out in going through some of the arguments.

The first question about this theory is why should we accept it? There are at least three reasons to be given. The first is economy. I take it to be fairly clear that quasi-realism would use slender means to explain a seemingly richer reality. Secondly, supervenience can be better explained in terms of quasi realism than in terms of realism proper. This needs some explaining. We often say about goodness: A thing x is good or bad in virtue of something. In the case of truth we are saying that there is something corresponding to the terms we use and in the case of goodness we are saying that there is a connection between the natural properties of an object and its moral properties. This seems to be the intuitive conception of supervenience. But the correspondence between a statement

15. Blackburn, S.: op. cit. p. 14

and conditions in the world is different from the correspondence between the two kinds of properties, natural and moral.

So how should supervenience be described? Are we saying that the moral property, goodness for example, is identical with the property in virtue of which it exists or upon which it supervenes? This is not what we say and it is also false. The goodness of an apple and the goodness of a knife are not identical nor are the natural qualities of apples and knifes upon which goodness supervenes in each case identical. So how should we describe this relation between evaluative and natural properties? How is good connected with other features of things? Is it a relation of necessity? The necessity in question cannot be analytic necessity because it does not make sense to argue that "good" means the same as some term of natural property, "sharp" for example or "sweet-tasting".

Blackburn describes the supervenience relation thus: it is necessarily true that if any particular type of things has both moral and natural properties, then all things of this particular type have them. Notice that this description of the supervenience relation does not say that it is necessarily true that if all things of a particular type have moral properties then they must all have natural properties. If this latter were true we could have an analysis of moral concepts in terms of natural ones. But this is specifically what supervenience does not supply. Some things could have natural

properties only and other things could have moral properties only. And we can imagine that the moral could float free of the natural. But supervenience tells us that it is necessary that if moral properties ever accompany natural ones they always do so. What this account of the supervenience relation excludes is that we could find things with moral and natural properties and other things with the same natural properties but not the moral ones in the same world. This is an impossible world according to this description of the supervenience relation.

The natural idea behind supervenience is that if a change occurs in the moral properties it is necessary that a change has occurred in the natural properties. The claim is that it is easier for quasi-realism to explain this relation than for realism.

The third reason for favouring quasi-realism is that it preserves the belief-desire theory of action. It is assumed that the belief-desire theory is superior to any other theory of the explanation of action. What this claim amounts to is not quite clear. But it seems to exclude the truth of internalism about motives for action. I have already pointed out that moral realism can either adopt an internalist or an externalist account of motivation for action.[16]

There are two more developments of this theory which need to be mentioned to make the description of it

16. Blackburn, S.: Spreading The Word, pp. 181-189 For an alternative see T. Nagel: The Possibility of Altruism, Princeton, Princeton University Press, 1970.

complete for my purposes. One is that it is a metaethical theory, not a normative one, a second order theory not a first order one. Two is that quasi-realism can take account of the fact that statements with value terms can occur in unasserted contexts which has been thought to count against all subjective theories of the meaning of value terms. It seems to me that formally such a subjective theory could be developed. If there are any troubles with it, they are to be found in its foundations, in the philosophical problems that seem to beset quasi-realism.[17]

These five points all tell in favour of quasi-realism. Quasi-realism is the latest development in subjectivism and it seems to answer most, if not all, the problems that have persistently plagued subjectivist theories of value.

The first point about economy is not really telling. The reason is that the economy of a theory is not necessarily confined to the fewest assumptions. If the overall shape of the theory is simpler, even though the assumptions are more numerous than of another theory,

---

17. See G. F. Schuler: "Modus Ponens and Moral Realism" in Ethics, Vol. 98, April 1988, p. 492-500, who argues that nothing like quasi-realism could possibly succeed. See also S. Blackburn's reply in the same issue "Attitudes and Contents" p. 501-517 where he develops his theory about indirect contexts in more detail and tries to explain how evaluative terms can occur in unasserted contexts. It seems to me that Blackburn can develop a formal theory of the occurrence of evaluative terms in unasserted discourse. Whether it ultimately succeeds depends on his analysis of the notion of logical form, validity and truth. To deal with this issue properly I would have to look closely at these notions. But I do not intend to do that.

then the number of assumptions is not a decisive point. Moral realists say, for example, that even though they assume moral properties this results in a theory of moral experience which is overall more unified and gives a better account of the moral world than a theory like quasi-realism which avoids positing moral properties. Considerations of economy are not obviously in quasi-realism's favour because it is not just the starting point that is important but the end-result as well.

The second point was that quasi-realism was in a better position to explain supervenience than moral realism. To see whether this is true we need to have a better look at supervenience and what is so problematic about it. When we say that an action is despicable or evil we are uttering a moral judgement referring to a moral value. Moral judgement is principled in the sense that any act meeting the same description would have to be described as despicable or evil. Supervenience captures the fact that moral judgement is principled and that we cite natural facts as reasons for moral judgements. Hence supervenience seems to be a seminal feature of moral judgement.

The problem of supervenience is the ban on mixed worlds. The ban on mixed worlds is the consequence of the description of the supervenience claim: It excludes the possibility of having one world in which a set of things of the same kind having some natural properties and some moral property and another set of things of the same kind having identical natural properties but no moral property

or a different one from the original set. This is impossible if supervenience is correctly attributed to moral values. If anything has moral and natural properties, then something else with identical natural properties must have the same moral properties. If the moral properties are different, then the difference must be reflected in differences in the natural properties. The impossibility is to have two identical things in natural properties with different moral properties in the same world. So either these two things with identical natural properties have no moral properties or they have the same moral properties. It is impossible for one to have and for another not to have the same moral properties.

Let us look at an example. Let us assume that abortion is murder from conception onwards. There are two women with relevantly identical background, they come from broken homes, started to drink early, have had many unsatisfactory love affairs. They both suffer rapes from their sexual partners. Is it possible that abortion in one case is permissible but in the other not? The fact that they are two separate persons does not affect the issue, unless their characters were significantly different. The assumption is that they are not. It does not seem to be rationally possible that one abortion would be permissible and the other impermissible. It is obviously possible that in one world both these abortions might be permissible and in another impermissible. But we cannot have that in one case the abortion was permissible

and in the other not in the same world. Our thought simply does not allow this; it is incoherent.

How should this ban be explained? The explanation of the ban is supposed to be a problem for realism but not for quasi-realism. The problem for realism is that it takes supervenience to be a relation between two kinds of properties of objects and there does not seem to be anything in the world which could explain the fact of supervenience. Quasi-realism claims to be in a position to explain this with features of our interests. Quasi-realism claims that it can grant all the ordinary claims of moral thought and yet that it should be explained by the notion of moral sensibility. If we allowed that we could respond to the moral qualities of two things identical in natural qualities in different ways we would divorce morality from its purpose. The purpose of it is to guide our actions and if it is possible coherently to respond in different ways to two things identical in natural qualities, then moral value is disconnected from practical decision making. So moral sensibility gives rise to moral qualities which in turn guide decisions.

The conclusion seems to be that the ban on mixed worlds is not a result of supervenience but of the fact that moral values require moral thought to be principled.

Now, let us turn to the distinction between first and second order questions in ethics, between normative ethics and metaethics. We mentioned this before as the

fourth point in favour of quasi realism. This will be the last argument to be looked at in favour of quasi-realism.

It is only fair to point out that this distinction has most often been used against realism as if realists could not accept it without seriously undermining their theory. At an intuitive level it seems to be straightforward that quasi-realists need this distinction but realists proper not because quasi-realists have to explain away certain features of ordinary language and ordinary moral experience which realists do not have to do. But it seems to me that there is no a priori reason for realists to reject the distinction. They could just as easily accept it and say that their theory was the better one because it made better sense of our ordinary moral experience than quasi-realism or because ultimately realism was a coherent doctrine which quasi realism was not. This would all presuppose that realism was a second order doctrine with no implications for first order views. Then both quasi-realists and other realists would believe that the theoretical debate between them would not make one jot of difference between the moral views they held. They might just as well accept either of these metaphysical theories and go on believing in democracy, the efficiency of the free market, the benevolence of the state, the sinfulness of humanity or whatever. No first order view would be a compelling reason for rejecting or accepting either of these doctrines. This seems to account for the view that this debate has no significance for ethics, its conclusions are neither here nor there

for rationally conducted debate and choice in ethics. This would amount to accepting the So What Thesis "that across a broad range of issues in and about morality it makes no difference whether moral realism is true."[18] I think it is only true to say that moral realists have usually taken the view that the distinction between first order and second order views must be rejected. In fact I know of no realist who has taken the view that this distinction is valid. But there are projectivists or anti-realists who want to reject it too and think it untenable.[19] It seems to me that ultimately the distinction is untenable as I will attempt to show below.

What is the status of this distinction? We describe it like this: we can hold any first order beliefs about morality without any implications or consequences for second order theories. Typical examples of first order views are that abortion is murder or capital punishment an effective form of punishment or that war is unjust. Typical examples of second order beliefs are that the truth or falsity of moral statements should be explained in terms of moral sensibility, or that there are moral properties in the world or that moral properties are intrinsically motivating.

It seems fairly clear that we can construct examples which straddle this distinction, where we progress from a

18. Sturgeon, N.: "What Difference Does It Make Whether Moral Realism Is True?" in The Southern Journal of Philosophy Vol. XXIV, 1986, Supplement, p. 115

19. This is what emotivists or expressionists similar to A. J. Ayer do. And this is what S. A. Rasmussen does in "Quasi-Realism and Mind-Dependence", in Philosophical Quarterly, 1985, pp. 185-191

first order view to a second order one without any obvious break. That's what examples like "All goodness is pleasure" seem to do. Another example could be that in ordinary morality we accept a principle that an act, even though bad, might not be blameworthy because of ignorance, i. e. if the agent did not know that the act was bad he is not as blameworthy as if he had known and if he could not possibly have known then he is not blameworthy at all. On the face of it this seems to be a second order principle with serious first-order consequences. The rejoinder to these examples is that there is no problem in talking about moral knowledge in ordinary first-order moral thought as long as we are careful that this does not imply anything about a second order theory about moral knowledge. These examples could be multiplied but this will have to suffice.[20]

The question that needs to be asked is: How should we understand the distinction between first and second order-questions, normative and metaethics? And if it is true that these examples demonstrate natural progression from one to the other is the distinction at all clear? The answer to this question is crucial for quasi realism because if the distinction proves to be unclear or unjustified quasi-realism seems either to collapse into old style subjectivism and not to be able to accept ordinary moral thought at face value or must find some further argument for this distinction.

---

20. See Sturgeon, N.: op. cit. where he discusses several such examples.

It seems to me that the distinction is not thought to be an inductive truth, i. e. a truth we discover from experience and infer from it to a general rule, in this case a distinction. If it were an inductive truth the examples I have described and others would throw serious doubt on it. The other option is that it is a deductive truth. But if it is, it has never been argued for in that manner, as far as I know. What is usually done is to point out that we can draw this distinction and we can pursue inquiries into the nature of moral values without committing us to any particular view of the practical issues. And this seems to me to be correct; it is possible to draw this distinction. But the fact is that there are links between first-order views and second-order ones as the above mentioned examples indicate. But if the distinction were deductively established, it would give the supporter of it a strong device. I have not discovered any such deductive argument for the validity of the distinction.

There seems to be a grain of truth in the distinction. The grain of truth is this: Any second order theory seems to be compatible with many, if not most, first order views. But it is not true to say that it is compatible with all such views. It is a commonplace observance that a consequentialist and a rights theorist can have identical first order views, e. g. that abortion is never justified but their arguments for these views would be very different. In general it seems to be true that second order views can make certain first order

views more or less plausible. A consequentialist for example is more likely to believe that abortion is justified under certain circumstances because the rightness or wrongness of acts depends on their consequences and these can change from one situation to another. But it is a perfectly possible position for him to say that they are never justified because the consequences of such a ban are preferable to abortions being allowed. Obviously the truth of such a view would depend on the truth of the factual claims about the consequences. If a rights-theorist takes the view that the foetus is a being with no rights, then abortions should be up to the parents or the mother because she is the only person concerned who indubitably has rights. But if he takes the view that it has rights, even a right to life from conception onwards, then abortion becomes questionable except in limited circumstances such as when the mother's right to life clashes with the right of the foetus. It seems to me it would be a curious view to hold a first order theory denying rights to any individuals, even persons, and a second order theory analyzing rightness and wrongness, goodness and badness, in terms of rights. But this is perfectly possible if it is true that any second order view is compatible with any first order view and there are no connections between any first order questions and any second order questions.

It seems to me that the onus of justifying the distinction is on those who want to hold it in this extreme form. It seems intuitively to contain a truth but

the examples mentioned above indicate that there are connections between normative ethics and metaethics, first order questions and second order questions. These connections do not invalidate the distinction as such but only in the extreme form I have described. It is a legitimate distinction. Quasi-realism needs the strong form of the distinction.

One reply to this might be to inquire further into the nature of the distinction itself. There seem to be at least three possibilities of understanding it. The first is that it is similar to the distinction between ethics and the philosophy of ethics. Philosophy of ethics can be practised independently of ethics but it is still a part of ethics. The second is that it should be interpreted along the lines of the distinction between any metatheory and what it is a theory about like the distinction between metamathematics and mathematics. Metamathematics can be conducted independently of mathematics but it is about mathematics. The third is the distinction between metaphysics and ethics. Metaphysics is not a part of ethics but a theory about the status of ethical statements and an explanation of how they are possible. Metaphysics is not a part of ethics and can be carried on independently of ethics.

It seems to me that Blackburn would want to understand the difference as the difference between metaphysics and ethics. He says: "So far I have discussed the metaphysics as if it were exclusively a second-order issue, with no necessary consequences for first order

moral theory."[21] This is quite explicit in saying that his projectivist theory is a metaphysical theory and equating the second-order and the metaphysical.

If this is Blackburn's understanding of the relation between first- and second-order questions then the question is whether it matters for what I have argued up to now. Blackburn can, it seems, hold on to his sharp distinction and say that all the examples I have mentioned can be considered in two senses: either they are first-order examples or second-order ones. Mackie's theory of error is flawed because it crosses the dividing line between these two kinds of questions. But there is one more argument to be considered about this.

If we see someone flaying a cat alive, the natural response is horror because doing such a thing is so vile. Anything so naturally, morally repellent has the moral quality of being evil. Quasi-realism wants to hold that such acts are evil whatever we feel about them just like realism proper. So quasi-realism like realism is committed to the falsity of this conditional:

(a) If our sentiments had been suitably different, then it would have been right to flay cats alive.

This conditional is a first-order moral statement.[22] Someone saying or uttering this conditional could be

---

21. Blackburn, S. : "Error and the Phenomenology of Value", p. 19
22. The argument here is taken from S. A. Rasmussen: op. cit.

taken as arguing for changing our sentiments or he could be regretting that this possibility is open. Now we might ask how this conditional could be justified and one way of making out how it could be is to use this more general one:

> (b) Had our sentiments been appropriately different, then moral truth would have been different, too.

It is not obvious that this conditional is true because it is ambiguous. If we want the right kind of connection to (a) then (b) ought to say:

> (b') Had our sentiments been appropriately different, then moral truth values would have been different, too.

Given this qualification it entails (a). But then we need another one to capture Blackburn's thought about moral truth.

> (b'') If moral truth is constituted by the best sentiments and the best sentiments can vary, then moral truth values can change.

Given this qualification, it seems that (b) entails (a). But as I have noted quasi-realism agrees with ordinary morality in believing that (a) is a false statement and would corrupt morality, were it true. Flaying a cat alive

is evil whatever we think or believe about it and however our sentiments are constituted.

As described above quasi-realism wants to accept the claims of ordinary morality and construe a metaethical theory explaining such claims in terms of our sensibility, the idea being that even though moral qualities strike us as being objective they are ultimately explained by the constitution of our nature. This can be done through defining the notion of moral truth by the notion of sensibility or attitudes, moral truth being the best set of attitudes at any particular time. There might be various problems in doing this, but I want to grant that it be possible. One obvious consequence of this theory is that it should accept the truth of (b) because if moral truth simply is the best set of attitudes then, if that set changes, moral truth changes. In this sense moral truth is mind-dependent. And because (a) seems to be entailed by (b) its truth seems to follow. But the heart of quasi-realism is the acceptance of ordinary moral thought and the moral objectivity that seems to go with it and at the same time have a second-order theory explaining this mind dependence and construing the concept of moral truth through the notion of the best set of attitudes and arguing that this machinery does all the work that is required from a moral theory. So there is a problem here.

It is perfectly possible at this point for Blackburn to dig his heels in and say these conditionals could be understood in two senses: they could be taken as first-

order statements or they could be taken as second-order statements and no connection between these two senses. There are two replies to this response. The first is that by doing it Blackburn seems to be guilty of equivocation. According to this response (a) would be false as a first-order statement but true as a second-order statement. But how is the first-order sense to insulated from the second-order sense? The second response is that the second-order doctrine does not seem very interesting and there do not seem to be many things to be said about it. The second-order sense seems to require some sort of quietism.

I want to consider one further argument about the validity of quasi-realism. Let us accept that the quasi-realism project succeeds. What we would have is a theory explaining moral truth in terms of our sensibility, a theory saying that in reality there are no moral qualities. Yet it wants to accept that ordinary moral thought is objective, assumes the reality of moral qualities and that they are independent of our awareness of them. One consequence of the project succeeding seems to be that no one should really think like a quasi-realist, because if we did that we would stop acting as if moral qualities were objective features of the world, that flaying a cat alive was evil whatever we thought of it or believed about it. It would not be independent of our psychological make-up. But this would defeat the premiss of the theory that it could accept moral thought as practised and as we have described it. One response to

this problem would be to say that this only demonstrated that the conditions for the theory being true were different from the conditions for accepting it as true. We might want to say that the theory was true but one consequence of it being true was that it could not rationally be believed. But this is impossible in terms of quasi-realism because moral truth consisted in the best set of attitudes or the most admirable sensibility. It follows that it is impossible to distinguish sharply between these two sets of conditions. If moral truth simply is the best set of attitudes we must be able to accept it. So quasi-realism cannot use the distinction between conditions of truth and conditions of acceptance.

These arguments seem to lead to serious doubts about quasi-realism as described. But the way out for the quasi-realist is to accept either anti-realism of some sort and try to get around the problems facing such a view and I have gone through. I do not want to say that this is impossible. To be able to do that I would have to tackle some difficult problems about the concept of truth for example. The other option is to accept moral realism.

1.4. Conclusions

In this first chapter I have tried to give some outlines of an objective conception of moral properties. I have also attempted to show how moral realism would reply to two rival theories about the relation between evaluation and description and about the analysis of

morality in terms of attitudes. It seems that moral realism has a satisfactory answer to these problems.

Chapter 2

Moral Realism and Value

In this chapter I will discuss four versions of moral realism. I shall reject them but try to accommodate the valid points in two of them in what I want to say. I will categorize moral particularism as a version of realism which depends on a particular conception of supervenience. I shall argue for an objective view of moral value. I shall also argue that it is reasonable to accept a consequentialist conception of the good.

2.1. Three Types of Moral Realism

There seem to me to be three ways of arguing for moral realism prevalent in the literature. First, the analogy with secondary qualities is used as a ground for saying that moral qualities are real and objective in the same sense as secondary qualities. This is weak moral realism. Second, moral qualities are thought to be analogous to primary qualities and can profitably be treated as qualities of objects. This is strong moral realism.[1] The moral realism I want to accept can

1. See J. Dancy: "Two Conceptions of Moral Realism" Proceedings of The Aristotelian Society, Supplementary Volume, LX, 1986, p. 167-169

usefully be thought of as a version of strong realism but without any commitment to an analogy between primary properties and moral properties. Third, moral properties are argued for on the basis of what we can call pragmatism. This means that the truth of our beliefs about moral properties is the best explanation of our moral observations and responses.[2] I want to keep this as a separate category because the pragmatic argument seems to me to be different from arguments for the other two positions. The fourth position, ethical particularism, is hardly a different type of moral realism. It is only discussed because it offers a distinct, and in my view mistaken, view of one feature of moral properties, i. e. generality.

It seems to me that most of the arguments, perhaps all, for moral realism can be put into one of those categories. This does not mean that there is a neat division between these three approaches. The literature on moral realism has grown so vast that no complete and exhaustive categorization is practical. How would dispositional theories of value, for example, be categorized?[3] This characterization should only be taken as an indication of three ways of arguing for moral realism. But to make the theory of moral realism clearer

---

2. R. Werner: "Ethical Realism" in Ethics, Vol. 93, July 1983, p. 653-679. "Moral facts exist, in the sense that I am using the phrase, just in case the most reasonable explanation of reports of moral observations include the positing of the existence of the moral entities mentioned in the reports." p. 653

3. See "Dispositional Theories of Value I, II, and III" in Proceedings of The Aristotelian Society Supplementary Volume, 1989, LXIII, pp. 89-174

we need to go through some of the arguments that have been offered and could be offered.

It has been offered as a strong argument for weak realism that there is a close analogy between moral qualities and secondary qualities. Ordinarily, the distinction between primary and secondary qualities is that primary qualities such as size, shape, weight are qualities of objects independently of anyone perceiving them, in so far as human beings can be said to perceive primary qualities. Anyone should be able to recognize them independently of the nature of their sensory apparatus. Secondary qualities such as colours are not independent of the sensory apparatus perceiving them. This means that the property of being red simply is the tendency to be perceived as red by a human being. The property of being square is the property of having four corners of 90°. It seems fairly clear that the nature of secondary qualities is dependent upon the sensory apparatus but the nature of primary qualities is not because this is in part simply how they are described. If human senses were different from what they are, the nature of the quality red would change but it would not affect primary qualities.

One inference drawn from this difference between these two types of qualities is that secondary qualities are not real in the sense that they are not objective. But this is not a warranted conclusion. Secondary qualities are just as real as primary qualities even though their nature consists in the possibility of being

perceived. But secondary qualities are remarkable in that they combine being subjective and objective. Being red consists in the tendency to produce a particular type of colour experience.

There are some things to notice about this description of secondary qualities. First, secondary qualities only affect those who have the appropriate sensory equipment. The same then presumably applies to moral qualities. The notion of "appropriate sensory equipment" implies the notion of normal sensory equipment. Second, moral qualities become anthropocentric in the sense that they are constituted by the possibility of a characteristic human response to them. Third, moral realism of any sort says that we form moral beliefs through recognition of moral qualities and they in turn generate responses. It then depends on whether the moral realist is an internalist or an externalist about motivation what relation is believed to obtain between belief, desire and action.

There are certain standard objections to weak moral realism which must be dealt with.[4] The first objection is that secondary qualities are differently associated with underlying physical differences from moral qualities. Differences in colour are as a matter of fact associated with underlying physical differences. But somebody who believes that two situations require two

---

4. I use C. Wright's summary of these objections in "Moral Values, Projections and Secondary Qualities" in The Proceedings of The Aristotelian Society, Supplementary Volume LXII, 1988, p. 5-11

different moral responses is obliged to believe that there are associated non-moral differences. But this should not worry us too much, because it is open to the proponent of weak moral realism to say that moral qualities only affect morally appropriate subjects and it is just part and parcel of being a morally appropriate subject to be able to discern non-moral differences associated with the moral differences. To this extent there is a difference between secondary qualities and moral qualities.

The second objection is that moral belief and moral concern are not the same. This means that it is possible to have two societies with identical moral beliefs but different moral concerns. But this is impossible on any conception of moral properties because it is an intrinsic feature of moral qualities that they are reason giving. The response to this problem is simply to point out that acquiring moral beliefs is acquiring moral concerns.

The third objection is that it invites unacceptable relativism into ethics. A moral quality does not only depend on what is in the object but also on how the subject perceives it. If the perceiving subject changes then moral qualities change. But this point ignores the normality built into the notion of an appropriately perceiving subject which puts constraints on how the subject can be. Moral qualities have their distinctive effects on a certain type of subjects. If the subject radically changes the qualities cannot have their

distinctive effects, but that does not mean that they have changed.

The fourth objection is that the analogy with secondary qualities cannot make any sense of being in extensive error about moral qualities, nor can it make any sense of widespread moral disagreement. The reason is that the quality of being red is constituted by normal human experience and response when it is registered in consciousness. If we want to hold that a population has fallen into extensive error about colour ascriptions, we lose hold of the concept of normal experience. It can be interpreted statistically in which case it would change when an extensive error arose; and it can be interpreted substantially or functionally meaning that it was the function of the sensory system to register their proper objects whatever they are. On this latter interpretation extensive error becomes problematic; we cannot hold that perceiving red, for example, is the function of the eyes, only that perceiving colour is their function. But a typical pair of eyes register red at a particular frequency of light waves. If this fact changed and most human eyes started to register some new colour, grue for instance, then we could say that they were flawed because this has not been their normal response. But this is a flaw only if the consequences of the changes are bad. If the changes do not have any bad consequences, this hardly counts as a flaw; the eyes would be more limited than before in not being able to perceive red. The question is whether the inability to perceive red counts as extensive

error, if there are no bad consequences. And when does a flaw constitute a new norm? It seems we would have to use the statistical notion to make sense of the error. But this notion cannot explain extensive error: As soon as error becomes extensive enough, it becomes a new norm.

The response to this objection is that in both cases the norm is set by our best responses in any particular case. In the case of perceiving red the best response would under normal circumstances be the typical response. But in the moral case it is not so sure that the best response would be the typical response. It is a commonly observed fact about moral life that it is difficult and the obvious symptom of this difficulty is that our typical responses often diverge from our best responses. This difference enables weak moral realism to make sense of moral disagreements between cultures and widespread moral disagreement or at least it opens up a space where sense can be made of these facts.

Weak moral realism can handle these objections, it seems, without too much trouble. But there is one more objection which needs to be considered. The analogy between secondary qualities and values requires that values have a phenomenal or perceptual aspect. What does this amount to? One suggestion might be this: perceiving rightness simply is judging something to be right. But the analogy seems to go further in the sense that there should be a distinctive phenomenal aspect to moral experience. But the problem with this proposal is that seeing a cat flayed alive for example might or might not

create a particular experience. But morality consists in being concerned for the right kind of thing for the right kind of reason, in this case concern for a living being because of the pain caused to it.[5] It seems to be true of secondary qualities experience that we can talk about raw experience of that sort meaning that ascribing hearing a sound to a child when it starts before it has any conceptual resources to identify such experience. In our reaction to the flaying of the cat nothing in that experience seems to be independent of the concepts we bring to bear on it. This seems to imply that moral responses are the formation of beliefs with moral content rather than the having of a certain sort of experience.[6] If this is correct weak moral realism seems to be in trouble because there does not seem to be any particular kind of moral experience.

It is a fairly well established fact that babies tend to cry when presented with parts of a face deformed or in the wrong relation to each other. Presumably they are afraid when they see it even though they have no conceptual resources to identify a face or any parts of it nor can they be conscious of the mental state of being afraid. I think most parents would accept that from fairly early on children start responding to pain in

---

5. Compare this part from Julian Barnes' novel Staring at The Sun: "...It may be different for others. I can't tell you what it's like being brave. You can't pick it up and look at it. When it's there you don't feel it's there. You don't feel excited or dizzy or something. Maybe you feel a bit more as if you know what you're doing; but that's all."(p. 47)
6. C. Wright: op. cit. p. 11-13

others for example. They easily cry and become afraid when somebody else is in pain. The question is whether such experience could count as moral or raw moral experience corresponding to the raw secondary qualities experience. But these responses can hardly count as moral until the child is in possession of the relevant concepts. Being afraid is not an experience of a secondary quality but it seems to have close resemblance to moral experience. It does not open up any space for raw moral experience. The same seems to apply to crying because of somebody else's pain.

But the point of these objections is that the intuition about the analogy between moral qualities and secondary qualities is that it can survive giving up the notion of moral experience. So even though the analogy disappears or gives out at this point the intuition can survive.

But ultimately it breaks down, if Wright is to be believed.[7] The reason it breaks down is that our best judgements about red, for example, determine the extension of the predicate "...is red". But our best judgements about evil, for instance, do not determine the extension of the predicate "...is evil". Ultimately our best judgements determine the extension of colour predicates but they do not determine the extension of moral predicates. The reason for that is that some of the conditions for applying moral predicates are based on a prior determination of moral status, determination, for

7. C. Wright: op. cit. p. 22-24

example, of whether the subject delivering the judgement is an appropriate moral subject. This shows that the mix of the subjective and objective is different in the case of moral qualities from secondary qualities. In other words our best responses in the case of secondary qualities can be characterized independently of the notion of redness but our best responses in morality cannot be characterized independently of the notion of morality; it is a moral question what our best responses are. So secondary qualities cannot serve as a good analogy for moral qualities.

This result should not really be surprising. Moral realism does not rule out the possibility of at least some subjectivity of the moral properties. When explaining actions by moral properties we could say that certain situations merited some particular response. The merits of any situation just seem more naturally placed in the objective sphere rather than the subjective one. In explaining why maltreating children is wrong no significant contribution is done by the fact that perceiving such a fact as a spectator can be painful. It is the pain of the child that is one of the relevant factors.

Strong moral realism holds that moral properties are analogous to primary qualities. This view holds that moral properties can be described independently of human responses and they are part of the fabric of the world waiting to be experienced. This view immediately runs into a problem. Primary qualities serve to explain events

in nature and have a causal function. Moral qualities serve to explain human actions but such explanations do not seem to be typically causal explanations. If you lodge a complaint against youngsters setting fire to a stray dog then the badness of the act is your reason for your action in a typical case. But it is not straightforward to say that the badness of the act is the cause of your act. This would mean that the pain of the dog causes you to try to prevent the action. But then the question arises why should the pain make us do this rather than encourage such actions. The reason seems to be that the painfulness of pain is bad. The badness of pain is not on par with the pain itself. The moral quality is needed to explain our action.

But moral properties are problematic and many who are willing to accept realism about science, for example, think that the notion of moral properties makes no sense and reference to them should be avoided in a good moral theory. One way to avoid moral properties in explaining actions is to argue that it is our belief that an act is wrong which explains our response rather than the wrongness of the action itself. If this explanation works, then moral qualities seem to become redundant. The crucial issue here is that it is the moral belief that is doing the work, not the moral quality. If it proved possible to construct an example showing that moral qualities can have effects independently of being recognized, then this would throw doubt, at least on this denial of any role for moral properties.

The Stalinist system of Soviet Russia has recently collapsed and a new order is slowly emerging. The swiftness of the collapse has taken many people by surprise. One obvious explanation of this collapse is to point to the gross injustice of the Stalinist system. This would be a moral explanation of a social event. It is open to a theorist to deny this explanation and say that moral beliefs are the decisive factor rather than moral properties. But it certainly seems possible to construct an explanation saying that the injustice of the system caused a lingering, unreflective resentment of it which in turn caused some of those living within the system to sympathize with its victims. In time this became fairly widespread and some people started reflecting on the injustices of the system and formed beliefs about it. When such beliefs were expressed they might have had explosive effects, if the resentment were widespread enough. If the expression of such beliefs was repressed, the system itself became morally bankrupt and in danger of losing the allegiance of the populace. If this is a reasonable account of possible events, then the prediction becomes plausible that when the repression would be lifted, the system would collapse.

An opponent of moral properties would say, at this point, that this example did not prove the point. The reason he would offer would be that the notion of resentment was an intentional notion and intentional notions and intentional states and concepts were problematic parts of a causal chain. The reason for this

is that intentional states have objects conceived of under certain descriptions. It is this fact about intentional states that captures moral properties. The injustice of the Stalinist system is this social system seen under a certain description. This state is of a different kind from one billiard ball hitting another and causing it to move. So it is the belief itself or its occurrence that enters the causal chain, not the moral property.

The question now is whether it is possible to make sense of the belief without reference to the moral property. But I do not want to pursue this argument further. I only want to claim that an explanation in terms of moral properties is possible but I admit that this does not refute the the possibility of explanation in terms of belief. I also accept that intentional states are problematic entities in a causal chain. But if moral explanation is possible, as I have described it, it does not seem to be radically dissimilar from other respectable causal explanations. The moral explanations would have to fulfill counterfactual conditions like other causal explanations, such as if the injustice had been absent, the resentment would not have occurred nor would the system have collapsed from this cause. This obviously needs some further argument, but it seems to be reasonable to accept that moral explanations based on moral properties could possibly be analogous to causal

explanations.[8] So strong moral realism seems not to be obviously absurd and capable reasonably to accept that moral properties might play a role analogous to a causal role in the explanation of social, moral and psychological events.

There are two main arguments which have been put forward for this strong version of moral realism.[9] The first argument is that taking our moral experience at face value we judge it to be an experience of moral qualities of agents and actions. And if we have no reason to think otherwise we take it to be indicative of how actions and agents are constituted. So seeing an innocent man condemned to prison we take it to be an unjust decision. This property of the act is not thought to be something which we in some way attribute to the act without there being anything corresponding to it. The second argument looks at moral choice. In making a moral choice we are not trying to get ourselves to accept any answer or to find an answer compatible with other choices we have made. We are trying to find the right choice and we have to present this search to ourselves as governed by a criterion which does not lie in ourselves. If that were not so how could the choice be wrong?

These two arguments might lead one to think that moral realism is the right way to account for our moral

---

8. See N. Sturgeon: "Moral Explanations" in G. Sayre-McCord: Essays on Moral Realism, New York, Cornell University Press, 1988 pp. 229-255 and Brink, D.O. op. cit. pp. 182-197
9. J. Dancy: "Two Conceptions of Moral Realism" in Proceedings of The Aristotelian Society, Supplementary Volume, LX, 1986, p. 172-175

experiences. The claim is that this is how moral properties strike moral agents in their moral deliberations. But it needs to be pointed out that these two points do not establish moral realism, they indicate that moral realism could be a plausible way of describing and explaining morality and our moral experiences.

But why should we not say that weak moral realism suffices for these two arguments? The weak version seems to capture the important features of our moral experience or moral thought and be able to account for the fact that such thought seems to be guided by external reality because secondary qualities are qualities of external objects. But as I said earlier the mix of the subjective and the objective is not the same in secondary qualities and moral qualities, so the analogy ultimately breaks down.

The third major way of arguing for moral realism is on pragmatic grounds.[10] The idea is to infer to the best explanation of reports of moral observations and experience. If moral qualities figure in the best explanation of our moral experience then we must assume that moral qualities exist. It must be admitted that in using the notion of "best explanation" pragmatic considerations have been used to argue both for and against moral realism. The notion of best explanation is spelt out as the one that is the most conservative, simple, general explanation of the fact at issue. As we

10. See Werner, R. : op. cit. Maybe White, M.: What Is and Ought to Be Done, New York, Oxford University Press, 1981 should be included.

saw in discussing quasi-realism this type of considerations lends itself to arguments for a more austere theory than moral realism, but realists do have a reply. This sort of pragmatism can be derived from Quinean holism which considers all statements to be connected and of the same kind as a complex web, meaning that there is no sharp distinction between analytic and synthetic statements. All statements are revisable but those statements which can be considered at the fringes of the whole web are more easily revised than those at the center such as the law of the excluded middle. But in principle all are. It seems to follow from this that explanations of experience are not necessarily constrained by the reference of statements expressing our evaluations. In explaining the truth of moral statements there is no guarantee that moral qualities would figure in the best explanation, even though apparently they were being referred to in such statements. So an explanation of reports of moral experience in terms of the observer's psychological set can count as pragmatic and then we would enter into similar considerations as earlier on about moral explanations in terms of moral properties and moral beliefs. As noticed then, explanations in terms of moral properties rather than moral beliefs are certainly possible and, it seems, sometimes plausible.

Another fact which underpins the pragmatic argument is that our experience is ambiguous. It seems to be true that any part of experience can be interpreted in different ways. This is what the "seeing-as" examples

show. It is just a result of our similar upbringing that we come to experience the world in the same or very similar ways. This should not be explained in terms of universal human nature or a unified structure of the world. If this is true it is just an interesting social fact that we tend to disagree more about morals than about science. It has nothing to do with the nature of the enterprises themselves or with the nature of the facts these enterprises deal with. The pressure to agree in science is greater than in morals and this is just a social fact about our post-industrial society but has no further metaphysical significance.

This type of pragmatism poses some difficult questions. One is about limits to theory formation. How does experience limit the possibilities of forming theories? Is there an unlimited number of theories which is compatible with experience? I will not answer these questions but accepting this theory would require an answer to them. I will not accept this theory. But it raises major questions about coherence, simplicity and other virtues of theories. In discussing moral knowledge in chapter 3 I shall try to accommodate concerns similar to those that appear in this type of pragmatism.

Moral realism, as I have described it and argued for it in the first and second chapters, is a metaphysical theory about the nature of moral properties. It says that moral properties are constituted by physical entities and that the moral supervenes on the physical but is not identical with it. It is possible to distinguish between

prudential and moral goodness, the difference between these two being that prudential goodness is goodness for me but moral goodness is goodness simpliciter interpreted here as agent neutral. This means that if relieving the wretchedness in the world affects A's good then it is a reason for anyone to do so. Moral properties make a proposition true or false and moral realism takes the view that moral propositions can be either true or false and that moral propositions cannot all be false.

2.2. Ethical Particularism

One view of moral properties subtly challenges what has been argued up till now about moral realism. This view can claim to be one type of moral realism but it denies the generality of moral concepts.

Ethical particularism is the view that there are no principles in ethics and our only option is to articulate the particular situations we find ourselves in at any point in time. The reason for saying that there are no principles in ethics is that moral concepts are not universal or universalizable. As should be immediately clear this is a radical view to take of the nature of ethics and an important one, if true.

There seem to me generally to be two ways of arguing for this view. First, I shall look at an argument for the view that the principle of universalizability is idle.[11]

11. P. Winch: "The Universalizability of Moral Judgement" in Ethics and Action, London, Routledge and Kegan Paul, 1972, p. 169

Second, a distinction can be drawn between universalizability and supervenience and the valuable features of supervenience can be kept but universalizability rejected. If these two arguments were true, universal morality would be false. I shall try to answer them.

The first argument goes something like this. There is a distinction between an agent and a spectator when considering any action. This corresponds to the distinction between first and third person statements. The principle of universalizability says that any person in relevantly exactly similar situations should think that" x is good" and "I ought to do x" are true. But persons are parts of the situations and they make a difference to whether two situations are relevantly identical or not, or exactly similar or not. It is also true that it is not practically possible to know everything about another person's actual or concrete situation. It is especially difficult to know how somebody sees the moral features of his situation. As spectators we can express a universal judgement on a particular action. But if it is true what we have said about the practical impossibility of knowing an agent's actual situation, then any inference from the universal judgement of the spectator to the agent's situation is blocked. Hence the universalizability principle is idle.[12] Another possible way of putting this would be to assume that every individual person is unique; therefore

12. P. Winch: op. cit. p. 169

any universal judgement about any action of any person is only applicable to that person. In this case the universalizability of moral judgement would be of no consequence for anyone. If an agent explained his action or inaction, why x was morally impossible for him for example, then he could go through the reasoning that led him to the conclusion. Another agent who came to an opposite conclusion would recount his reasons for his action. There would be nothing more to say than observe that different moral possibilities presented themselves to these two agents.

If this is correct, then the principle of universalizability is indeed idle. But this seems problematic. There is nothing wrong in observing that our knowledge of others is limited, as indeed it is. Anyone who has ever tried to fathom another person in depth, must have discovered that somehow something always escapes. The understanding never seems to be complete. But this is just humility which any moral theorist should acknowledge. Does it really limit universalizability as was claimed?

The second argument was that when supervenience and universalizability were properly distinguished the latter would lose all plausibility as a feature of moral thought. Hare says in one place that the thesis of universalizability is composed of two things: universality in the sense of the exclusion of references

to individuals and supervenience.[13] This shows that Hare at least thinks that supervenience is a part of universalizability.

There is no need for an extensive explanation of supervenience here. The basic idea is that a supervenient quality is consequent upon some other property; in the central case for us a moral quality is consequential upon natural facts. Described like this supervenience is an ontological quality. But supervenience can be described as a quality of moral judgement. Then it says that moral judgement must be consequent upon moral facts. This is how Hare considers supervenience and this is how we must consider it in this context. Universalizability is the requirement that moral judgements are dependent only upon universal facts.[14] When phrased like this it seems reasonable to think that universalizability is part of supervenience rather than the other way around as Hare thinks. How should we think of the relation between universalizability and supervenience?

The central idea in supervenience is that moral judgements are consequential upon facts: X is judged good in virtue of some fact, social, economic or natural. Any particular moral judgement can be supervenient upon any non-moral fact, any number of natural qualities or even upon the whole of natural qualities in an action as described.

---

13. R.M. Hare: "Supervenience" in Essays in Ethical Theory, Oxford, Clarendon Press, 1989, p. 74
14. See S. Blackburn: Spreading the Word, Oxford, Clarendon Press, 1984, p. 220

Hare thinks of this relation as one between a particular moral judgement and all the natural properties of the situation or object in question. At one place he is discussing two pictures P and Q and either P is a replica of Q or vice versa. "Now there is one thing that we cannot say; we cannot say 'P is exactly like Q in all respects save this one, that P is a good picture and Q not.'"[15] The force of the "cannot" in the quotation is that to allow divergent judgements about the goodness of the picture is contradictory. So supervenience seems to bring universality along from the start. This is the explanation for Hare's sometimes expressing the universalizability in practically the same terms as supervenience. "The thesis of universalizability requires that if we make any moral judgement about this situation, we must be prepared to make it about any of the other precisely similar situations."[16] It is as if universalizability were supervenience by another name. But the explanation for this is that the nature of supervenience brings in universalizability. This seems to me to be Hare's reason for thinking of the relation between supervenience and universalizability in the way he does.

But this seems clearly false. Wit seems to be based on a supervenient judgement. It is true about a witticism

---

15. R.M. Hare: The Language of Morals, Oxford, Clarendon Press, 1952, p. 81, 131. See also Freedom and Reason, Oxford, Clarendon Press, 1963, p. 18-19 and "Supervenience" in op. cit.
16. R.M. Hare: Moral Thinking, Oxford, Clarendon Press, 1981, p. 42

that we can respond to it differently on similar occasions without anyone finding it curious or blameworthy. When we hear a joke for the second time the wit can be wearing thin. This makes judgement about wit different from moral judgement. But both are supervenient. One inference seems to be obvious from this: Universality or universalizability is not part of supervenience but supervenience is a part of universalizability. Establishing supervenience does not lead to acceptance of universalizability.

Another observation seems to be in order. Hare's example of the pictures seems more readily to be assimilated to wit than to moral properties. Let us ask 'Why can't we possibly say about the two identical pictures, P and Q, that one is good and the other is bad?' The answer is that we can. If one is a replica of the other, then it is not good in the same sense as the other one. It is only if we assume that the two pictures are exactly alike and made by two talented artists independently of each other, that we will be ready to say that they are both good. But this is an extremely unlikely possibility. If we are talking about the more likely possibility that one is a replica of the other, then we have no reason to say that they are both good. The question of origin is crucial in attributing goodness to works of art. To this extent, at least, the supervenience of aesthetic judgement is more similar to judgement of wit than to moral judgement.

The set of facts moral judgements supervene upon is not the set of all non-moral facts. Some non-moral facts just are not relevant to the moral judgement in question. But how should we characterize the set of non-moral facts that are morally relevant? One obvious way of doing it is to say that they are the facts in virtue of which we judge an action as good for instance. The idea is then that whenever these facts are present a judgement that it is a good action results. This combines supervenience and universalizability in a way congenial to Hare as described above. But there are two problems with this explanation of supervenience. One, it is not possible to have a list of facts which guarantees a particular moral judgement whenever these qualities occur. The reason is that moral judgements are such that they can be defeated in particular circumstances.[17] I might want to say that a girl should be judged as good in virtue of being chaste, pious and considerate. The implication seems to be that whenever a girl is chaste, pious and considerate she ought to be judged good. But what if a girl, having all these attributes, is also cruel? Should we say she is good? No, and the reason is that the attribute of cruelty defeats the goodness. But is the answer to this not that among the reasons for calling the girl good originally is that she was not cruel. The answer is negative. "I do not call someone good because he is <u>not</u> cruel, though I may refrain from calling someone good on the grounds that he

---

17. J. Dancy: "On Moral Properties" in <u>Mind</u>, 1981, Vol. XC, p. 376-377

is cruel."[18] Two, as a result of this feature of moral judgements it is not possible to infer from particular instances of moral judgements to any general criteria for the ascription of moral judgements. This rejection of the possibility of any general criteria is fully compatible with moral judgements being supervenient in a more limited sense than that suggested above. The girl is certainly judged good or evil, as the case may be, in virtue of something: She is judged good because she is pious, chaste and considerate. But this carries no further implications for any other relevantly similar person. Three, the number of non-moral facts is indefinite and there is no subset of non-moral facts always giving rise to a moral judgement of goodness, for example.

A direct result of this argument against supervenience and universalizability seems to be ethical particularism, the view that each and every situation must be articulated without any comparison with other situations and that there are no general principles to back up the descriptions. In any particular situation the agent chooses some particular feature of it as especially salient and there is no thought of it being generally salient.[19] This is similar to aesthetic appreciation. In trying to appreciate a picture or a building we describe it and in describing it some prominent features of it are mentioned as especially important within the context of

18. J. Dancy: op. cit. p. 377
19. J. Dancy: "Ethical Particularism and Morally Relevant Properties" in Mind, 1983, Vol. XCII, p. 546

the whole object. This mention does not carry any general implication for appreciation of buildings or pictures. Picking a particular action out as wrong is not to give evidence that it is wrong but to try to get your audience to see that particular action as wrong and to see it that way is to join in the judgement of it being wrong.

This is most certainly a powerful challenge to the universality of moral judgement. If true, there are no general, moral principles possible and no universal moral judgements.

The first challenge to universalizability was that the principle of universalizability is idle. The reason was that when it was formulated from a spectator's point of view in the 3rd person propositions no inference could be drawn about the personal point of view or 1st person propositions. The reason is that it is not possible to say anything about how the moral possibilities appear to a particular person on the ground of some general 3rd person statements.

In thinking about moral matters we are often wondering about how we ourselves would have behaved. So we make hypothetical judgements about ourselves in hypothetical situations. This is perfectly compatible with Winch's argument because he does not want to deny that the principle of universalizability applies to our judgements about our own conduct at different times in different situations and in hypothetical situations. It also applies to our judgements of the conduct of others. To all these kinds of judgements the principle of

universalizability applies. But it does not follow from this that we can say that others will come to the same conclusion about those situations as we have done or that they should come to the same conclusion.

The question here is whether we should accept the universalizability thesis in this weak form. It is fairly obvious that in the example Winch describes in detail, from Melville's story Billy Budd, that no other position is reasonable. But this example is problematic. It is a literary example and literary examples are closed in an important sense.[20] So the appropriate judgement about Captain Vere's action seems to flow from the description of the example. Moral judgements are based on descriptions of actions and in real life it is usually contentious how actions should be described.

Another reason for doubting this limit on the universalizability principle is that there seem to be fairly clear examples where we do not hesitate to say that others ought to judge a particular action in the same way as we do. In saying that causing unnecessary suffering is wrong we do not just think this to be coherent with our previous judgements or with our hypothetical judgments or that it is wrong for us. We are judging it to have the moral property of wrongness. It is obvious that we could be mistaken about this property, but if we are not then this is something which every moral agent could and should discover himself. Winch

20. O'Neil, O.: "The Power of Example" in her book Constructions of Reason, Cambridge, Cambridge University Press, 1989, pp. 172-176.

denies that he is arguing for a relational interpretation of moral properties, i.e. that A is the morally right thing to do amounts to saying that A is the morally right thing for me to do.[21] Nor does he want to be understood as arguing that if G thinks that A is the right thing for him to do then A is the right thing for him to do. Winch thinks that the difficulty here is that we are trying to bridge the gap between propositions and expressions of decisions.[22] The question is whether this is really possible. His suggestion is that deciding what to do is an integral part of finding out what is the right thing to do. It does not make sense to say that I can decide what somebody else will do. Hence, I cannot find out what is morally right for others because part of finding out is deciding what they will do.

It is perfectly true that it does not make sense to say that I can decide what someone else will do. But it makes perfect sense to say that I can decide or find out what is morally right for someone else to do. This is not the same as deciding what he will do. If he accepts my decision about what is morally right for him to do, he is likely to do it. The reason that it makes sense to say that I can decide or find out what is morally right for somebody else is that I can know what is morally right. Usually when I have bothered to find out what is morally right I do not need to decide what to do because I know what I ought to do. If I am my normal self I do what I

---

21. Winch, P.: op. cit. p. 164
22. Winch, P.: op. cit. p. 165

ought to do. I can admit that my knowledge of others is limited but this does not mean that I am completely ignorant of their beliefs and desires. In certain circumstances I can even be in a better position than somebody else I know to judge what is good for him and what is morally right for him to do taking into account his beliefs and desires. What is morally right for somebody to do is not something mysterious and inaccessible to everyone else but him. Admittedly, people can have notions about right and wrong that are very different from our own and such differences make all discussions about moral rightness and wrongness difficult, sometimes even impossible. In the most difficult cases it can be difficult to find common ground, but in some cases these different notions can actually be worse than our own, permitting actions which ought not to be permissible. It is possible to compare different systems of values and we very often do. This implies that there is something which is morally right and it is not necessarily a part of any particular moral system as we know it. The difficulty of knowing other people and the fact that inferences from propositions in the 3rd person to propositions in the first are not obvious and straightforward, does not mean that there is nothing which is morally right and is accessible to everyone.

The basic reason for this being so is that morality is not just personal but social. Morality is a social institution bringing along a code of conduct accessible

to all those who are part of that particular morality. If it were possible to say that the moral principles were not the same for everyone, then this social aspect of morality would be weakened, if not destroyed. Justice is the supreme example of a social principle forming a major part of any morality. It can be extremely complex and difficult to apply the principles of justice but the basic principles are accessible to anyone and apply to all.

There is another thing which should be considered in this context. Hare often says that universal principles correctly applied fully determine the right action in every conceivable circumstance. This they do in virtue of their universalizability. But practical judgement plays a part in applying the moral principles and it cannot be guided by principles in doing that. If it were, we would need still other principles for those guiding the practical judgement in application. These would be second order principles. Then we would need third order principles to guide us in applying the second order principles and so on. This would be an infinite regress. So the application by practical judgement of moral principles is necessarily not guided by moral principles or other principles. This does not mean that it is random, though.

So it seems to me that this much is true in ethical particularism: there is an ineliminable role of practical judgement in moral appreciation or understanding. We should also point out that moral situations do not come

fully carved up to confront our recognitional capacities. They must be described and to describe them we use terms, sometimes moral terms. Such descriptions must be justified. It may be true that ideology always plays some role in what is considered to be a justified description of a situation. But we are not completely free in describing any situation as we wish. If that were true, we would neither need nor be able to justify such a description.[23] In justifying a description of a situation we must refer to reasons and reasons seem to be general in form.

It is open to an ethical particularist to say that moral properties are like the salient features of a building and are not essentially tied to a comparison with anything else. This is how we should think of reasons we give for our descriptions. But it should be noticed that the examples used are of artistic objects, a picture and a building. There is nothing inappropriate in thinking of aesthetic properties in this way, as not essentially comparable to properties of other aesthetic objects. I know we often compare the aesthetic qualities of a work of fiction and a building to the aesthetic qualities of other artistic objects but the most remarkable thing about them is their differences and that we are not obliged to think that another relevantly similar work is good. There is no comparable option open

---

23. On these points see: O. O'Neil: "Abstraction, Idealization and Ideology in Ethics" in J.D.G. Evans: Moral Philosophy and Contemporary Problems, Cambridge, Cambridge University Press, 1987, p. 55-69,

about moral objects. There is nothing obviously absurd about holding that it is of the essence of artistic objects that they make many such interpretations possible. But we are very reluctant to accept this about moral qualities. The explanation seems to be that a moral judgement is principled in a sense that an aesthetic judgement is not. It should not be inferred from this that an aesthetic judgement is less objective than an ethical judgement. Properties can be objective without requiring the judgement about them to be principled. I do not want to say anything about that. This fact of an ethical judgement being principled I have described earlier as being concerned to do the right thing for the right reason. In doing this we come to see a certain situation as a situation of a particular sort and that this sort of situation requires an action of this type. The moral concepts used by our practical judgement make us see it this way or that by constructing descriptions of the situation. These descriptions are not unique and they must be justified. In justifying them we give reasons and reasons are general in nature. Moral reasons based on moral values seem to be general in the sense that, in so far as they are moral reasons, they are reasons for any moral agent.

2.3. Reason, Value and Desire

We do not seem to be willing to accept that there can be two or more opposing views about a moral action

that are correct. Moral judgement is principled in the sense that we should do the right thing for the right reason. Implied in this description is strong resistance to the idea that two contradictory or opposing views of a moral action can both be correct. When considering whether two opposing views are possible of theft or murder, for instance, we are not primarily thinking about how particular instances should be described, although this is clearly part of our considerations. What we are most interested in is whether theft or murder described as such can properly be accepted as both right and wrong.

One way to explain this reluctance would be to think that what is right does not depend on our desires or on our practices. This brings in a number of considerations on subjectivism and objectivism about value, on truth as important for moral values and on the place of reason in moral thinking.

One way of looking at the nature of values is to say that they consist in the obtaining of certain psychological states. An action can be good if it produces pleasure, for example, or bad if it produces pain. On this view the goodness of an action consists in the psychological state it produces. A desire-satisfaction theory of values falls into this category. It holds that goodness consists in the satisfaction of a desire and badness in frustrating one. This is an instance of a subjective theory of value because satisfying a desire gets its value from the psychological state accompanying it, the satisfaction. Satisfying a

desire for a glass of water means that I have to fill it with water, put it to my lips and drink the water. But this activity in and of itself has no value, if the desire-satisfaction theory is correct. It is only the satisfaction of the desire which gives value to the activity. The satisfaction of a desire is a psychological state because desire is a psychological state. So desire-satisfaction theories are subjective theories.

It could be objected to this that a desire satisfaction need not be a psychological state. What should we say about a desire for x when x is in the future? This means that the desire can be satisfied when x occurs, even though I shall never know that it has occurred. This is not really a denial of the psychological theory. What it denies is that the satisfaction must be a conscious mental state. It does not deny it is the satisfaction of desire that explains value. But this might also be taken to imply that value could be explained by reference to interests or the good of somebody. Even though this could not explain all possible value it is certainly true for much of what we deem valuable.

The question is whether desire-satisfaction theory of value can account for our thought and talk about value. One fairly obvious implication of the desire-satisfaction theory is that the value of certain actions, traits of character or things depends on their ability to satisfy our desires. So things are good because they can satisfy desires. I believe that in ordinary moral thought

we take it to be the other way around: We desire things which are good. This is not a fatal flaw in the doctrine. If it is a part of a fully fledged subjective theory of morality, it can defend itself by distinguishing between first and second order questions. I think this is implausible, as I have tried to argue before.

But because value depends on the satisfaction of desire, if this theory is true, the possibility opens up that we could have the desire satisfied in different ways from those that are most frequent and natural. This new way could be much easier and more effective than those we usually rely on. Drinking water from a glass you hold to your lips can, on occasion, be fraught with difficulties. If our psychological states impregnate our actions with value, then, perhaps, we should try to bring about the psychological states in the simplest, most effective way possible. Nozick's celebrated experience machine might come in handy.

> "Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to

> your brain. Should you plug into this machine for life, preprogramming your life's experiences?"[24]

The obvious answer is no and Nozick gives reasons for a negative answer. The basic reason is that we actually want to be a certain kind of person, do a certain kind of thing, to really affect the world. If this is correct, then the view that actions receive their value from the experience they cause or the psychological states they cause is false. So the desire satisfaction theory seems to be false.

This is not the end of the story because desire-satisfaction theories and objectivism about values are not contradictories. It is certainly possible to have a theory mixing objective and subjective sources of value. I will give three reasons to think that objectivism is a plausible account of values.

The first is that we can distinguish between two kinds of desires. Some desires are value-laden and some desires are not value-laden. Value-laden desires are dependent upon evaluations. If I want to be a good father, it is because I believe it is valuable and desirable. I believe that bringing up a child is a valuable kind of activity. If I want to be a writer, it is because I believe writing to be a valuable kind of activity. Desires like these do not confer value upon those activities; they presuppose the belief that the

---

24. Nozick, R.: Anarchy, State, and Utopia, Oxford, Basil Blackwell, 1974, p. 42

activity desired is valuable. It is taken to be valuable independently of the desire for it.

Non-value-laden desires do not presuppose evaluations. An example of a non-value-laden desire can be my desire for a chocolate bar. In explaining my desire for a chocolate bar I do not have to presuppose a belief of mine that the chocolate bar is valuable. It is just that I want a chocolate bar as I sometimes want a glass of water. These desires have to compete with others to issue in action. But if we wanted an explanation of why we want a bar of chocolate, it seems to be sufficient to say that eating it leads to a pleasant mental state. And this seems to be all there is to say about this desire. But as will become clear I accept that all desires seem to presuppose desirability characteristics possessed by their objects.

It could be argued by someone believing the desire-satisfaction theory of values that the value-laden desires could be reduced to the non-value-laden ones. This seems to be the only option open to him because he cannot accept that there are actions, traits of character or objects that are valuable independently of their being objects of actual or counterfactual desires. I do not believe such a reduction can be carried through, but let us accept this for the sake of the argument.

There seem to me two replies to this move. The first is the objection implicit in the experience machine. If it is correct to reject the offer of plugging in for a planned out lifetime of pleasure, then this seems to

amount to a rejection of any desire-satisfaction explanations of value. Accepting a desire-satisfaction explanation of value seems to come to accepting the offer to plug into the experience machine.

The second reply is this. Even when explaining desires that are non-value-laden we seem to be able to give a cognitive explanation so that the desire as such becomes redundant. A cognitive explanation is an explanation in terms of how the agent conceives his action which seems to me ultimately to come down to what the agent believes or conceives to be true about the world.[25] If we say that instead of a bar of chocolate somebody desires to consume a saucer of mud, then we are certain to be puzzled. It really is no explanation to say that he desires the saucer of mud and believes that the saucer contains mud. The problem is that the desire for mud is unintelligible and we just cannot see how anyone could find mud attractive. What is needed to explain this strange desire is to try to see how the agent views the desired state of affairs, i.e. what he takes to be true of his circumstances, his world. This seems to me to amount to reducing non-value-laden desires to value-laden desires or better to reduce non-value-laden desires to beliefs about the desirability or value of objects, actions or traits of characters.

The second reason for accepting objectivism about values is that desire-satisfaction theories seem to give greater value to activities and lives which we ordinarily

25. See McNaughton, D.: op. cit. pp. 110-113

do not think to be valuable. The problem is with two kinds of desires which, if satisfied, do not seem to increase value in the world.[26] The first kind is immoral desires. A committed Stalinist wants to put his opponents into jail, to internal exile or before an execution squad. In exterminating his opponents he satisfies his most important desires. He is willing to justify his desires by the overriding importance of bringing about a classless society. Such desires are obviously immoral because they completely disregard the interests of others and their satisfaction will decrease value in the world rather than increase it. But this is something a desire-satisfaction theory of value has difficulty in dealing with. The second kind is desires that are trivial. A desire to be able to fly in the air meditating in the lotus position is a fairly clear example. The satisfaction of such a desire can obviously be important for the one who has the desire. But it cannot reasonably be said to be valuable. Its satisfaction will not increase value in the world, unless we take the view that satisfaction of a desire is increasing value. But by accepting that according to our ordinary notion of triviality this desire is trivial it seems to follow that the satisfaction of this desire does not and cannot be valuable. The point is that the desire-satisfaction theory cannot explain why this is so. It seems to be better explained by supposing that there are objective

26. See Brink, D.O.: Moral Realism and The Foundations of Ethics, Cambridge, Cambridge University Press, 1989, pp. 226-227

values. This does not logically follow from the failure but it certainly makes the thought that there are objective values explaining this failure of the desire-satisfaction theory more plausible.

The third reason for objectivism about value is the implied rejection of the instrumental view of reason in what has been said already about value. The instrumental view of reason states that reason is only concerned with how best to fulfill our present desires. It attempts to discover the best means to satisfy the ends supplied by our desires. Questions about the rationality of desires do not arise. This does not mean that all desires are completely indifferent to considerations of beliefs. The desire to go to the Scottish Opera or the desire to have a pint of beer at a particular pub can be discussed and they are subject to considerations such as whether it is possible to get a ticket for the Scottish Opera, whether you can reach it in time for the performance, or whether the pub is within reasonable distance and sells some good beer. But these desires are derived desires in a sense and we should distinguish between them and underived desires. The desire for a drink is the basic or underived desire on which the other desire depends, the desire to hear music is basic to the desire to go to the Scottish Opera. The claim of the instrumental theory of reason is only that these basic or underived desires are given and not subject to reason.

This claim is not implausible at first sight. But there are three reasons for doubting it. The first reason

is the untenability of the distinction between these two kinds of desires. I have already discussed the distinction between value-laden desires and non-value-laden desires. Ultimately it seemed that even with non-value laden desires we had to have an account of why we desired what we did in terms of the agent's beliefs. This becomes clearer when we look at the second reason for doubting the instrumental view of reason. The second reason is that some desires seem to be irrational or immoral. The desire to kill or exterminate one's political opponents or class enemies is clearly an immoral desire. A desire to destroy the world is on ordinary understanding suspiciously non-rational, maybe irrational. It is certainly not absurd to say that they are irrational or immoral. If the instrumental theory is to be believed this should be absurd. The third reason is that we seem to be able to discuss perfectly properly whether some desires are valuable. This certainly applies to the desires that are immoral or irrational. What about a desire to eat or a desire to drink? Somehow they seem to be such that questions about their value do not arise. But even though they do not actually arise there is nothing to imply that they could not arise. It is just that we find them so natural that questions about their value seem pointless. But as soon as we have strange desires like a desire for a saucer of mud, as we saw earlier, it should become clear that we can, in principle at least, have an answer to a question why we desire x which comes down to what we believe to be true about its

value. These three reasons are strong indications that the instrumental view of reason is false.

In arguing for the objectivity of value it seems to follow that the function of reason becomes wider than the instrumental view implies. Reason is involved in establishing the truth or falsity of beliefs and considerations of truth or falsity play a major role in finding out whether a desire is valuable or not. It seems to be possible in principle to ask questions about value of every desire. So desires do not carve out a part of reality that is immune to reason. It is true that desires can be stubborn and refuse to bow to rational considerations. But this is not surprising. A desire for a cigarette is irrational given the desire to live, if it is true that smoking causes lung cancer and, possibly, death. But this desire can be immune to any such considerations and can cause serious damage to one's health. The same applies to a desire for a bar of chocolate. Consuming considerable amounts of chocolate can make you fat and being fat can damage your health. We can certainly ask about the value of such a desire. What is not subject to rational considerations at some point in time is my liking of chocolate. The fact of my liking chocolate need not be based on any rational considerations. It is just something I find out about myself. But over time this can change and my liking can disappear in view of my rational belief that eating chocolate can damage my health. So when explaining my

wanting to avoid chocolate I must refer to my belief that chocolate is bad for me.

We do have particular beliefs and we do have sets of beliefs. I have a set of beliefs about a particular bar of chocolate, for example, that it is rectangular, brown, wrapped in paper and so on, and about chocolate in general. We also have beliefs about beliefs or second order beliefs. We have already seen how questions about the truth of particular beliefs arise. We can also ask questions about the truth of a set of beliefs. If we form more general beliefs we can have a theory about a particular domain of beliefs. We can apparently have a theory about moral beliefs, for example.

Moral particularism which I discussed in 2.2. denies that a moral theory is possible. One reason that might be given for this could be that moral properties supervene on natural properties and moral properties are multiply realizable. But this is not convincing. The fact that the property of being good can be realized in various types of behaviour has no bearing on the possibility of generalizing theory about moral properties. A denial of theory about morality must be based on a particular view of the nature of moral properties and their nature.

As I hoped to make clear in 2.2. moral particularism denies that moral properties can be general and one reason for this is that there is always the possibility of countervailing properties. I want to claim that moral theory can take into account that moral properties can be outweighed by other moral properties. It also seems to be

true that the principles of a moral theory can be as specific as we need and yet they can be universalizable. Specificity is opposed to generality but universalizability is opposed to singularity.

There are reasons for accepting theory in morality. One is that morality is a social institution that guides our actions or ought to guide our actions. I have had reason to accept this earlier. This is especially important when we are dealing with cases that are new and which our previous moral judgements have not covered. In such cases we are unsure how to act and what to believe. It is also important when our evaluative views are in conflict, either when our own views are in conflict or when they conflict with the views of someone else. In resolving conflicts we appeal to general principles or very general evaluative views. This relation between particular moral judgements and general principles is the essence of moral theory. We try to find general principles which explain our particular moral judgements and we test general principles and theories by seeing whether they have any counterintuitive consequences and whether they fit our moral judgements. If a theory has widespread counterintuitive consequences it loses plausibility, but if it has problems in explaining some limited number of our particular moral judgments, this need not be decisive against it. It is not only that our beliefs must fit the theory but the theory must fit the beliefs. So if some beliefs do not cohere with the theory, we might have to revise the theory. The result of

this process is a set of considered beliefs which are hopefully true and coherent. This set of considered beliefs is a result of applying the theory to our judgements. So if we believe we have a reason for action because we desire x but this desire depends on valuations that do not survive our reflection on these valuations, then this is not a good reason for an action. It does not justify the action. This shows again why the instrumental view of reason does not hold.

2.4. Consequentialism

I hope to have established the conclusion that moral properties are objective, that moral facts seem to have a role to play in explanations. I have also argued that values are objective. These conclusions are compatible with a number of substantive theories about the nature of moral values. They seem to be compatible both with deontology and consequentialism. In this last part of the chapter I want to make the thought plausible that consequentialism is a good view of the nature of moral values.

There are five reasons why I think consequentialism is plausible. 1. Impartiality seems to be a necessary feature of consequentialism. We can at least sometimes evaluate an action or an agent impartially. If we come to be in a position of a judge in a dispute between two agents, then we should try to come to a conclusion which is fair to both. This requires impartiality which I take

to mean that the judge takes into account the facts of the case without reference to who these agents are. He tries to come to a right decision. If he were a judge of a court of law he would have to take the law into his considerations. But this impartiality seems to require that we can think of the good of agents neutrally. We can think of several outcomes of the dispute and we can rank them according to a principle. This is not saying that there is only one such principle possible. Prima facie at least there are many principles possible for such a ranking.

This reason could be objected to on the grounds that impartiality created so serious problems for consequentialism that it should be thought a reason against it rather than for it.

This objection requires a threefold reply. First, the importance of impartiality should be explained in some more detail. Second, the problems impartiality creates for consequentialism need to be described. Third, it must be considered whether consequentialism can respond to and explain those problems.

Morality has to do with the good of rational agents. One common sense reason for accepting impartiality appears when we ask whether we should let anyone's interest count for more than the interest of another. It is at least prima facie plausible to accept the principle that the interests of one ought to count for no more than the interests of another. Another reason for accepting impartiality is the universalizability of morality. If it

is true that moral principles are valid for all it is plausible to infer that impartiality is a feature of morality. A third reason for thinking of impartiality as a central feature of morality is agent-neutrality. Impartiality directly follows from agent neutrality. I have accepted both agent neutrality and universalizability as features of moral principles. These reasons seem to explain why we take impartiality as a most important feature of morality.

The second part of the reply to the objection mentioned above is describing how impartiality causes problems for consequentialism. There are at least two ways for impartiality to cause problems for consequentialism. The first results from the feature of consequentialism which says that the best or right action in any circumstances depends on what the best consequences are. This can easily be thought to contradict the requirements of impartiality. The idea, presumably, is that maximizing the good can easily lead to the disregarding of someone's interest. But this would be a misunderstanding. Even if we accept that all good is the good of individuals and that impartiality is a central feature of morality, this does not have the consequence that ultimately no one's interests should be disregarded. The impartiality requirement is only that everyone's interest must be taken into account. This does not rule out that ultimately maximizing the good might lead to disregarding someone's interests. It might be better, impartially considered, that the interests of

someone were disregarded. The second way for impartiality to cause problems for consequentialism is that it seems to require us to disregard the personal point of view or our personal projects. But it is these personal projects that give value to our lives, are the sources of our deepest happiness. But if it is a fact that disregarding the personal point of view has worse consequences than not disregarding it, consequentialism tells us not to disregard it. It would not be maximizing the good.

The third part of this response is that consequentialism does seem to have replies to this worry about impartiality. This not to deny that other questions arise because of these replies. But there is no space to discuss them here.

2. Moral reasons tell us what we ought to do. In general it seems to be true that everyone ought to maximize the good and minimize evil. This seems to be a principle of practical reason. In conjunction with 1 this leads to the conclusion that we morally ought to increase goodness impartially.

3. Morality is a social institution enabling people to fulfill their needs and desires together. Morality is impartial. This seems to mean that the interests of each moral agent have equal weight. So in attempting to find the best action in a given situation one needs to judge how it affects everyone. This does not necessarily lead to the best state of affairs being that state which benefits everyone equally or harms everyone equally or leaves everyone equally unaffected. The best state of

affairs can easily be unequal. The fact that everyone's interests have the same weight only means that everyone's interests have to be taken into account in judging what the best state of affairs is. It is perfectly possible that someone's interests must be disregarded if the best state of affairs is to obtain. But this disregard must be taken into account when judging what the best state of affairs is.

4. One might argue against this formulation of consequentialism that one could not distinguish between the bad effects of accepting something as true and its falsity. On any plausible account of the matter this should be possible. But this is only an effective charge against consequentialism if it denies that moral statements can have truth value. If moral realism accompanies the consequentialism this will cause no problem. This is because moral realism enables us to distinguish between the truth conditions of a theory and the consequences of believing it. In the extreme case consequentialist moral realism can accept that the acceptance of a true theory could have such disastrous consequences that it had to be repressed. It can even explain how this can come about because it can distinguish between truth conditions and acceptance conditions. This last point shows how consequentialist moral realism can deal with this problem and by doing

that shows how it can be a true theory even though accepting it could have disastrous consequences.[27]

These reasons for consequentialism have not established it as true but hopefully they have given it some plausibility. Consequentialism is not obviously true. The reason is that consequentialism raises a number of difficult problems and seems to require a radical revision of our common sense moral views. One consequence of accepting that one should maximize good by aiming at the best state of affairs seems to be that the circumstances might arise when a rational agent had to be victimized. The standard example is to ask what one ought to do if faced with the choice between killing one by one's own hand or have terrorists kill ten. Consequentialism seems to require that it is morally right to kill one. But this conclusion is repugnant to our entrenched moral views.

Another problem is that consequentialism seems to require too much from us if we are to achieve the best possible state actually available in any situation. This means that I might be monstrously morally wrong in attending a course in philosophy because I could have gone to Ethiopia to save starving children. There are two responses to this. One is that it might actually have better consequences, if I did what I took to be important, rather than pursuing the course which appears

---

27. See Sturgeon, N.: "What Difference Does It Make whether Moral Realism Is True?" in The Southern Journal of Philosophy, Vol. XXIV, 1986, Supplement, p. 132, where he makes this last point.

impartially the best. The other is that any appreciation of the consequences of a course of action has to take into account a realistic human psychology. We might argue that it is not realistic to expect human beings to attempt to bring about those states of affairs which are impartially best.

This is not the place to enter into a full discussion of the merits of consequentialism but suffice it to say that at least some of these problems seem to be soluble by consequentialism while it itself probably needs refining.[28] But the best way of examining the theory is to see whether it can make sense of the questions that arise when tackling an important problem. This I shall do when considering paternalism. Paternalism seems to be a problem which is best accounted for in consequentialist terms.

There is one more consideration which seems to support consequentialist view of the nature of value. Persons are purposive beings. This means that they have ends and under normal circumstances they desire these ends and they act on the beliefs they have about the desirability of those ends. To say that persons are purposive beings implies that they do what they do either for some benefit they are attempting to secure or to avoid some harm or because they think their action right. This seems to mean that all intentional actions of

---

28. See for example the essays in S. Scheffler (ed.) Consequentialism and Its Critics, Oxford, Oxford University Press, 1988, esp. by T.M. Scanlon and P. Railton.

persons have some apparent good as their aim. This coheres well with the fact that it is apparently problematic to account for actions and desires that aim at the bad. But there are some clear instances of actions that do not aim at the good or even the apparent good which count as exceptions to this rule. There are incontinent actions, i. e. actions the agent performs and intends but does not have sufficient strength of will to perform. These actions show weakness of will on the part of the agent. It may also be that the agent's reflection has led him to the conclusion that some of his desires are not the desires he wants to have. So second-order desires are possible which are desires about first-order desires. This makes it possible that the motivational structure of an agent can be different from his conception of good. So actions caused by desires the agent does not find desirable are an exception to the general rule that intentional actions aim at an apparent good.

For actions successfully to achieve the good they set out to achieve they must be rationally ordered and coherent. This prevents conflict between desires but conflict can be damaging for action. Reason puts our desires in order either by putting priorities on them or by getting rid of desires that are contradictory. I assume that when we have found that two desires contradict each other one of them gives way. One of them must be irrational. But this is relatively rare, I believe, in comparison to putting priorities on desires.

Priorities resolve conflicts of desires. Desires can also be irrational in the sense that we are overcome by them and have no control over them. A desire to jump from the top of a cliff is an irrational desire of this type.

A third sense in which a desire is irrational is that it can involve a false belief. This can either be a belief about a fact or a belief about a value. I am thirsty and I believe that there is water in that glass. Hence I drink from it. Unknown to me someone has put transparent glue in it. It is a bit steep to call this desire irrational because the belief in question is false. What we usually need in addition for it to be irrational is that the belief resists correction or the desire for that glass is active after the discovery of the glue. But in this context I do not believe it makes much difference that I call such a desire irrational. There can also be irrational beliefs about values. I might desire the destruction of the whole world rather than move my finger. This desire is based on the belief that moving my finger is more desirable or worthwhile than the destruction of the whole world. This belief is irrational and the desire as well. In so far as a belief is irrational it prevents the action from achieving its aim. It is clear in the example of the glass of water. This seems also to apply to an irrational belief about values. If an agent acted on the belief about value described above, the destruction of the world would prevent him from moving his finger. But usually values are taken to be irrational because of other reasons.

Their irrationality then comes down to abnormality or unnaturalness. It is false to think that moving my finger is more worthwhile than destroying the whole world. Hence it is irrational in a similar sense to the belief about the water in the glass. For the agent to be rational and successful in his actions he must eliminate irrational desires and beliefs. This he does through reflection upon those beliefs and desires and through experience. It does not seem to be possible to know all the desires well without having experience of them, i. e. seeing how their consequences are in practice. This process creates considered moral beliefs which ought to be coherent under ideal circumstances.

Let us think about an agent who is contemplating whether to inject himself with heroin or not. He has the desire for the pleasurable state the narcotic induces and he realises the dangers to himself from injecting the drug into his body. This knowledge does not persuade him to resist his desire; he thinks the present desire and its satisfaction are more important to him than his future state of mind and judgement. This can indicate that he has not discovered his ends or his good.

It seems that there are at least two ways for reason to discover ends. One way of doing it is through reflection on the conduct of others. By thinking about their conduct and seeing imaginatively why they do what they do their aims may appeal to us. If we had no aim similar to that which motivates their conduct it can become ours. This does not only happen reflectively

though. These aims only become ours when we decide and do act on them. This is a more complicated process than one might think and there do seem to be all sorts of reasons stopping us in our tracks trying to improve ourselves. There is sheer laziness, indifference to what is admirable, and weakness of will and more numerous obstacles in our way.

Another way of discovering new aims for our conduct is not to glean it from the conduct of others but to discover it in ourselves, so to speak, to discover what we really want rather than what we apparently want. Our motivational structure is not clear enough to make fully transparent to every agent everything what he wants at any point in time. Some of his desires are clear to him but some are not.

Ends are not fully determinate and sometimes we do not know whether a particular action is a part of one or not. We know of clear instances of justice, for example, but it is sufficiently indeterminate to prevent us from saying whether another action which we have never performed before would be an instance of justice or not. The only way of finding out is to do it. In doing it we discover whether it is an instance of justice or not through our experience or perception of the moral qualities of the circumstances.[29]

In these considerations we have gone over some of the ground covered earlier and it seems to me to

29. See Dahl, N. O. : "Rational Desire" in Sartorius, R. ed. : Paternalism, Minneapolis, University of Minnesota Press, 1983, p. 265-266

reinforce some of the points I have already made. In particular it seems to me that this account of persons as purposive beings supplies one more reason to accept consequentialism as a theory of moral value. If persons aim at the good and more good is better than less, then we ought to maximize the good inherent in the ends we pursue.

2.5. Conclusion

In this chapter I have argued that moral properties are objective and that they can be parts of explanations of social phenomena. I have argued against the view that moral properties are particular. It seems to be plausible to accept that values are objective. This means that they do not depend on desires and we can say that a person has interests which are independent of his desires. Consequentialism is independent of the considerations of objectivity but it has some features that make it a plausible substantive doctrine about the nature of moral value. This gets further support from considerations about the nature of persons.

Chapter 3

Moral Knowledge

In this chapter I will argue that knowledge is possible and the best account to be given of the justification of knowledge is coherentism. I will also argue that moral knowledge is possible and we can be justified in claiming to know a moral property. The account of the justification of knowledge is unified in the sense that if it is correct for knowledge in general it is also correct for moral knowledge. Lastly I will argue that a counterexample to this theory, Williams' theory of moral knowledge, cannot answer some serious doubts. The conclusion will be that it is just as reasonable to suppose that moral knowledge is possible as knowledge is in general and that it should be explained in the same way.

3.1. Knowledge

'A believes that p' is a proposition that is usually taken to be relatively unproblematic, the relation of believing not being too demanding on related notions. The major condition for believing is that A understands the concepts he uses in the proposition. If he does not, there is a difficulty in identifying the belief in question. But there is no requirement that a belief be

true for it to be a belief. It is a natural supposition in all or most ordinary cases of believing that the persons take their beliefs to be true, but any belief that is false for whatever reason does not stop being a belief.

'A knows that p' expresses a different relation between A and that state of affairs referred to by 'p'. Knowledge is a different relation from believing. This can be seen from the fact that there is a requirement of truth built into the notion of knowledge. If we believe that we know something and discover later that 'p' in the proposition 'A knows that p' was false, we infer that there never was any knowledge. Knowledge requires truth and if there is no truth no knowledge is possible. It should also be born in mind that the minimum requirement of understanding applying to belief applies also to knowledge.

This can be expressed thus: A knows that p iff 1) p is true, 2) A believes that p. We must add the third condition: 3) A is justified in believing p. The third condition is reasonable. A belief can either be true or false and we must have evidence for taking it to be true rather than false. This evidence can, for example, be sensory. This is why we should add 3).

The problem, one might say, with this theory of knowledge is that justification is no guarantee of truth. And we need truth for knowledge. So even with the best of justifications there is always the possibility of a gap between knowledge and justification. This gap the sceptic

exploits. The problem is even more serious than that as the Gettier examples show.[1] Let us assume that there are two people in my house, a friend and a stranger. A knows that his friend has a Renault car because he has seen it and travelled in it but unknown to him the friend has now sold it and the stranger has a Renault car. It is true that someone in A's house has a Renault car and he justifiably believes that someone does but yet he does not know that someone in his house has a Renault car. This is a Gettier example. What it shows is that even if our belief is true and justified it does not amount to knowledge. Knowledge and the justification of knowledge are independent of each other.

This is a flaw in the traditional theory of knowledge. A number of responses are possible to this flaw. One might be to say that knowledge must be causal, so that the state of A and what p refers to must be causally linked. This immediately makes moral knowledge, the knowledge of the future and mathematical knowledge problematic. The basic problem with the causal theory of knowledge, though, is that it requires different explanations of the justification of true beliefs from the explanation of the justification of false beliefs. A true belief is caused and is justified in terms of the cause, but a false belief is without a cause and cannot

---

1. See E. Gettier: "Is Justified True Belief Knowledge?" in A. P. Griffiths (ed.): Knowledge and Belief, Oxford, Oxford University Press, 1967, p. 144-146. See also J. Dancy: Introduction to Contemporary Epistemology, Oxford, Blackwell, 1985, p. 23-36. These examples have been widely discussed.

be justified in the same way as a caused belief. This means that we cannot have the same kind of justification for true and false beliefs which is implausible. So other attempts have been made.

What needs to be accounted for, roughly speaking, is the non-accidental nature of the relation between knowledge and the thing known. The most prominent and most discussed recent attempt to account for this relation and its nature is Nozick's conditional theory of knowledge.[2]

Nozick accepts the first two parts of the analysis of knowledge mentioned above but instead of the third he puts two others which capture what is needed to avoid the Gettier examples and what is essential about knowledge according to him. These two additions come in the form of two conditional propositions: 4)If not p, then A does not believe that p; 5) If p, then A believes that p. In the case of the Gettier example described above, the theory can handle why this is not an instance of knowledge. The four propositions 1), 2), 4) and 5) must all be true, if A is to obtain knowledge. Proposition 4) is not true of the Gettier case. It would look like this, if it were true: If no one in A's house had a Renault car, A would not believe that someone did. Let us suppose that the stranger is absent or owns no car or a different kind of

---

2. R. Nozick: Philosophical Explanations, Cambridge, Massachusetts, The Belknap Press, 1981, 3rd chapter, p.167-288. What I say about the theory will be very brief and the important pages are 172-178. See also J. Dancy: op. cit. p. 37-49. There are all sorts of complications of Nozick's theory which he discusses into which there is no reason to go here.

car from the Renault. But in this case A would still believe that someone in his office owned a Renault car, that is to say his friend. This offends against our understanding of the conditional in question. The subjunctive conditional p--->q says "that in the situation that would obtain if p were true, q also would be true."[3] Correspondingly the conditional says that were p false q would also be false. This understanding of the conditional 4) has the consequence that it is not true of the Gettier case. Hence, A does not know that there is someone in his house that owns a Renault car.

This conditional theory captures the non-accidental nature of knowledge but it is not without its problems.[4] The most important of these is the problem of internalism. Nozick argues that his theory is an externalist one and that it is one of its most important virtues. To get a grip on this the distinction between internalism and externalism needs to be explained. Let us take an example. If A wakes up each morning with a belief about which horse will win the race that day and it turns out that A is right, then Nozick would accept that A knows which horse will win that day.[5] In such a case the subjunctive conditionals hold and A's belief tracks the truth. This can happen even though A has no idea how he knows. The distinction between knowing and guessing becomes blurred.

---

3. R. Nozick: op. cit. p. 173
4. See J. Dancy: op. cit. p. 44-48
5. R. Nozick: op. cit. p. 266

There are two things that need to be said about this. The first is that the key idea in externalism is that A can know p without believing or recognizing that he knows. Knowing without believing that one knows is on the externalist account a genuine instance of knowledge.[6] Because the relation of tracking is a factual relation in the world between the beliefs in question and the facts or truths, this consequence of the theory should not be surprising. Second order knowledge is analyzed in terms of the subjunctive conditionals just like first order knowledge. So knowing that one knows is just keeping track of one's knowledge that p. The second thing that needs to be said about this example is that the same question arises about justification as about knowledge itself because it is implied in the example that there is a reliable method through which knowledge is obtained. The question: how do you know that you know? is a question about justification.

It seems to me that the question about second order knowledge cannot be discussed without discussing justification. The major question here is whether it is plausible to think that knowledge is possible without second order knowledge or beliefs. I think everyone would hesitate, at first at least, to ascribe knowledge to A just on the basis of his uttering one morning which horse will win and turning out to be right. But given consistency in his utterances and predictions and that

---

6. R. Nozick discusses this on p. 245-247

the subjunctive conditionals apply, then the consequence of this theory is that A knows what horse will win.

It is not quite clear what sort of generality is implied in this account. Nozick thinks generality is implied: "...since tracking involves some generality to other (subjunctive) situations, it seems plausible that whenever a person knows, there will be some reliable submethod via which he knows."[7] So if we have the subjunctive conditionals turning out true in some particular circumstances of A and we can say that in relevantly similar circumstances he would know, then the theory is more plausible than before. It could also be argued that thinking that epistemology should supply us with an account of second order knowledge is unreasonable. All it could do is to explain first order knowledge and if such an account worked it would be an effective reply to the sceptic. To try to do more would be hopeless and ultimately gives the sceptic all the room he needs to manoeuvre to undermine knowledge. If we must add to the four propositions above another one 6) A believes that 4) and 5), as is plausible, if we are to include second order knowledge, then still another one must be added 7) A's belief that 6) is justified and yet another one 8) A's belief that 7) that A's belief that 6) is justified because it tracks the truth. This leads to an infinite regress and dooms internalism.

The obvious justification of internalism is the naturalness of the notion that knowledge necessarily

7. R. Nozick: p. 267

implies knowing that you know and, at least in favourable cases, knowing how you know. If you do not know that you know that p, then you do not know that p. What is troubling in the example of knowing which horse will win the races that day, is precisely that the agent is not able to give any account of that he knows or how he knows. In saying 'A knows that p' we can fill in any proposition we like for p. In the example of the horse we might put 'that Red Rose will win the 3.30pm race at Ascot today' or 'that grass is green' or 'that increased CO_2 in the air causes the heating of the atmosphere' or 'that genocide is morally wrong'. Obviously these are different beliefs. Now the question is this: Is it plausible for all these beliefs to say that A knows them if they are true and A's beliefs track the truth? It seems plausible to say that any rational agent would have different reasons for accepting these beliefs. But on the externalist account the bare truth of any belief and, as should be clear from the quote from Nozick, the reliability of the belief generating method, are sufficient to make it knowledge and to justify it as such. Now any externalist could accept that any rational agent had various reasons for accepting these different beliefs, but what it ultimately comes down to is that truth and reliability generate knowledge, even though the knower is incapable of recognizing his knowledge. But one thing should be noticed here. What is reliability in a knowledge generating process? The consistent transfer of truth from fact to belief. But this seems just another

way of saying that a belief tracks truth. So nothing is really added by using the notion of reliability.

Pointing to the truth of any belief is an unsatisfactory justification of it as knowledge. In the case of the horses we would suspect that A had a lucky guess rather than he knew that Red Rose would win, even though we could say that his belief tracked the truth repeatedly. It is plausible to justify empirical beliefs by pointing to their truth, beliefs like 'this patch of grass is green' and even with the general proposition 'grass is green'. But I do not know what it amounts to to point at the truth of 'genocide is morally wrong'. Causal knowledge as in 'increased carbon dioxide in the air causes the atmosphere to heat' is problematic as well. Showing that this is true is not just pointing at increased heat and increased carbon dioxide but showing why one should lead to the other. This obviously takes a lot of background facts for granted. The lesson from this is that the more complex the knowledge becomes the more implausible the externalist account becomes. The notion of justification cannot be exhausted by the notion of truth. Pointing to the truth of a belief is not sufficient for A to be justified in believing it.

So what is being justified in believing that p? In justifying we cite other beliefs and facts which are supposed to lend credence to p. The obvious question here is: what about these other beliefs and facts? What lends credence to them? If they still need other beliefs and facts, the same question arises about these facts and

beliefs. And so on ad infinitum. The consequence would be that there are no justified beliefs. The only other option, it seems, is to accept that there are beliefs where justification comes to an end. Such terminal or ultimate beliefs are not justified by anything else: they are self-evident or fundamental in some other sense. If such beliefs were found, we would have secured a basis for all our knowledge. There have been various candidates for such beliefs. One is beliefs about our present mental states and going along with that is the notion that we cannot be mistaken about our present mental states. Hence, they are a good candidate for self-evident beliefs. Another one could be some general and generally accepted beliefs about the world anybody takes for granted such as 'the world has existed for longer than 5 minutes' or 'I have two hands'. Propositions expressing this knowledge are more certain than any other propositions doubting them or adduced as a reason for them. If we have an argument saying that we do not know that we have two hands, for example, we can be sure that one of the premisses of such an argument is wrong or there is a mistake in the reasoning leading to the conclusion that we do not know. This argument about the certainty of knowledge or the lack of it is fully coherent with what I want to say about knowledge and moral knowledge, so I shall not say anything about it now. But the first possibility is more problematic than it seems. The most important characteristic of our present mental state is that it seems to be infallible.

If there are such states of which we can truly say that we infallibly know that p, they would probably be the best candidates for the ps that must be self-evident. What does this notion of infallibility come to? What is it that we cannot be mistaken about and how is it that we cannot?

Our present mental states can be described with propositions like 'This appearance of a red patch I seem to see before my eyes'. The words in this proposition express concepts and these concepts are true of a certain range of things. It seems possible that in applying these concepts to my present mental state I might make a mistake. But what does such a mistake amount to? One view could be that it is a mistake in applying the concept of 'red', for example, to the kind of colour appearing in my visual field. But this can be either of two things. It can be that I have made a mistake about the word 'red', my belief about the word can be mistaken; it does not express the concept I thought it did. So this is no mistake about my present mental state. This is no trouble for the infallibility view being considered here. But another understanding of the mistake is possible. It is not just that we can be mistaken about what concept a word expresses but we can also be mistaken about the things the concept applies to, in this case about the experienced colour appearing in my present mental state. Assuming that I am fully attentive and have deliberated about my words and concepts, then a mistake of this nature is a substantial mistake about the quality of my

experience. It may be that the concept which is true of the experienced colour of my present mental state is 'pink' rather than 'red'. This would then not just be a mistake about the word but a mistake in my understanding of my mental state and hence a mistake in my belief about my mental state. The consequence of this, it seems to me, is that no infallible beliefs are possible given that one's present mental state is the best candidate for a state about which we can be infallible.

There are other qualities of beliefs which might make us think that they could serve as foundational beliefs. Qualities like incorrigibility and indubitability. An incorrigible belief is a belief no one could ever be in a position to correct and an indubitable belief is a belief no one could ever have a reason to doubt. The problem with both these sorts of beliefs is that they do not make sense without infallibility. If it were possible to have incorrigible beliefs and these beliefs could possibly be wrong, this would be a profound disaster in our intellectual make-up. I do not see any reason for believing that rational beings are subject to such a disaster nor do I see any reason for arguing for this possibility. We would be better off without such beliefs.

The conclusion which should be drawn from this is that no basic or foundational beliefs seem to be possible. But this means that no convincing account has been given of knowledge yet.

Earlier I rejected the view that we could give an account of knowledge based on the principle that justifying beliefs need to be justified themselves because it led to an infinite regression. After having found the foundational view wanting, it is reasonable to have a closer look at the earlier view. The first question might be: why is the infinite regression so bad? Because if the view is true, knowledge is impossible. So if this way of arguing for knowledge is to get anywhere the regress must be stopped.

In discussing the possibility of mistakes about our present mental states one thing that might have been mentioned was that in using words to describe anything we are committing ourselves to describe like qualities alike. In saying that some thing is red we are committing ourselves to using that concept about any other thing appearing to us to have the same quality. This is a condition for the intelligibility of language in general. This seems to mean that comparison with other judgements is involved in any particular judgement. So in judging something to be red we are comparing it with other things which are red or which are not red. A belief that x is red is justified by other beliefs about the same type of quality and beliefs about similar qualities. So beliefs seem to be justified by other beliefs.

But this has in no way avoided the problem of infinite regress. It only indicates what we are doing when forming a belief. But if we are comparing beliefs and this was all that could reasonably be required from

us, when justifying, why should this infinite regress be bad? Maybe knowledge is possible compatibly with such regress. One suggestion might be this: if we are satisfied that the belief we are checking coheres with other beliefs and is not in direct contradiction to them, can't we say that we are justified in saying that we know that belief? It seems to flow from the fact that in forming a belief we are comparing, that we cannot check beliefs in isolation. Even though we can single out a particular belief and come to a conclusion about it whether it is true or not or whether it is justified or not, this has consequences for innumerable other beliefs. So even if we think we have a very secure belief, an empirical belief for example, this does not mean that it could not change in the light of other beliefs. In that sense our beliefs form an interconnected whole. If one belief is changed, many others will have to follow.

Beliefs can possibly be infinite in number but the chain of justification of our knowledge cannot be, because human beings are finite. Maybe infinite regress will not be a serious problem, but the question arises whether justification will be circular. A belief p is justified by belief q which in turn is justified by belief r but because this is not infinite it will ultimately have to come back to p. This seems to imply that p is ultimately justified by p which is viciously circular. This does not seem to be any better than the infinite regress.

A distinction can be drawn between two kinds of justification. Justification can be linear and it can be inferential, it can be either of these and it can be both.[8] The vicious circularity assumes that all justification must be linear. This means that p is justified by q, q by r, r by s and so on. Were this the only type of justification available, we would either have an infinite regress or vicious circularity.

But there is the inferential justification. And it has been claimed that it avoids vicious circularity because it allows "justificatory chains to loop back upon themselves."[9] The notion of inference is thought of here in the standard way as going from premisses to a conclusion. But this ultimately comes down to inferring p from p. So what is the difference between inferential justification and other kinds of justification? There are two possibilities here, at least. One is to say that inferential justification is always conditional in the sense that it is only if q and r are true that p is true. So we are not saying that p is true but p is true, if q and r are true. This means that any belief will only be true conditionally. This is no vice of such a theory. Coherentism should be understood in such a way that the justification of p is its contribution to the coherence of a whole set of beliefs. So the conditionality of any particular belief should not be surprising.[10] The

8. See D.O. Brink: Moral Realism and the Foundations of Ethics, Cambridge, Cambridge University Press, 1989, p. 105
9. D.O. Brink: op. cit. p. 105
10. See J. Dancy: op. cit. p. 55-57 and 127-130

justification of any particular belief, then, is its contribution to the coherence of the whole set of beliefs. If we accept this, the question to be answered is how do we justify sets of beliefs. Here it is possible to say either that a fully coherent set of beliefs is unconditionally and actually justified, and that this blocks regress, or that no justification of a coherent set is possible. Justification is defined in terms of contribution to coherence of belief sets, but this does not mean that justification and coherence are one and the same thing.

The other possibility of explaining knowledge in terms of inferential justification is to draw a distinction between a systematic and contextualist justification.[11] Contextualist justification is in essence incomplete justification. It applies the principle of justifying beliefs being justified to those beliefs being checked but assumes some beliefs to be known to be true even though not justified. A systematic justification does not leave any belief unjustified but applies the principle of justifying beliefs being justified to all beliefs. I take it to be obvious that such an application of the principle cannot possibly be linear. If the principle of the justification is applied to all beliefs, then there will always be one belief unjustified at least. But the idea in inferential justification is that ultimately justification is allowed to loop onto itself, as mentioned earlier. The question

11. D.O. Brink: op. cit. p. 123-125

is whether this is any better than saying that p justifies p, just in a bit long winded way? The answer is no, it is no better, unless something else is added. "...one's belief p is fully or systematically justified insofar as p is part of a maximally coherent system of beliefs and p's coherence at least partially explains why one holds p."[12] This way of explaining justification comes down to the same thing as discussed above. Justification as a contribution to coherence stops the regress and makes knowledge possible.

It should be noticed that having a systematically justified set of beliefs, as just described, cannot be a realistic prospect for human beings for the simple reason that beliefs are infinite but a human life is finite and chains of justification must be finite. It should also make us suspicious of this formulation of the justification of knowledge that the set of all truths or all true beliefs is a very problematic entity, to put it mildly, and this would be the set resulting from systematic justification. So systematic justification is a problematic notion. But this should not worry us too much. Giving up the notion of a fully systematic notion of all true beliefs does no damage to the coherentist view of knowledge here espoused. It only means that all our systems of beliefs are provisional in the sense that it is always possible to have another one more comprehensive and more coherent.

---

12. D.O. Brink: op. cit. p. 124

In the quotation above it said that p being a part of a coherent system of beliefs at least partially explained why one held p. This is a reasonable requirement. In so far as we are rational beings we must have a reason for accepting p to be true and this reason must justify our belief that p. If justification is thought of in terms of coherence, it follows that p being a part of a coherent system of beliefs should explain our belief that p, at least in part. This also means that having a reason to believe p is having second order beliefs about the belief that p. Such beliefs would be beliefs about why the belief that p came about, what type of p it was. I suggest that what was lacking in the example above of the man who knew what horse would win the race later that day was rationality. He could not give a reason for believing that he knew this and hence could not justify it. This seems to lead to internalism about justification and there does not seem to be any way of avoiding it. But it should be pointed out that it is possible for coherentism to be externalist because it seems to be possible for beliefs to be coherent without a rational agent being aware that they are.

One question needs to be answered about this view of knowledge. Is this coherentism compatible with the realistic view of moral properties taken earlier in this essay? It seems to be natural to think that a maximally coherent set of beliefs guarantees truth and knowledge and maximal coherence seems to be the best possible reason for truth and knowledge. This would mean that

truth and knowledge could not be evidence transcendent and a maximally coherent set of beliefs could not be false. Having attained it we would have everything we could possibly have as evidence for knowledge. But this is a denial of realism as I have described it.

There is a reply to this worry. First, let us recall that we rejected the notion of a set of all true beliefs. If there was one we might want to say at some stage that we had reached it or that at some stage in the future we could. This does not seem to be either a realistic or a reasonable hope. Why we should accept that we have at any point in time got to a maximally coherent set of beliefs of which it could not be said that another one was impossible which was more coherent is not clear. If there is no such point, then the possibility of being wrong opens up and, it seems, that truth could transcend evidence. Second, the beliefs we have about beliefs, i.e. the second order beliefs, are realist beliefs as should be clear from earlier chapters. These second order beliefs must cohere with the first order beliefs and these second order beliefs provide strong evidence that a maximally coherent set of beliefs is true. But, as should be clear, they do not provide a guarantee of truth. It is always possible that maximally coherent sets of beliefs are wrong.

One could question at this stage whether the relation between second-order beliefs and first-order beliefs was the same as between first-order beliefs and the world. We claim truth for our first-order beliefs, if

we succeed in knowing something, and evidence of truth for our second-order beliefs. This evidence consists in coherence with other second-order beliefs and first-order beliefs. But it will not do to claim that the truth of all first-order beliefs consists in coherence. The theory would then be a pure coherence theory.[13] The problem with pure coherence theories is that they lack grounding and we do not seem to have a reason for believing them.

I have anyway been willing to accept the possibility that even though our beliefs were justified they might be false. This means that I have to accept for some beliefs that they are not true in virtue of coherence. This means that the theory is an impure coherence theory and such theories combine coherence and correspondence. Their main problem is that they are using two concepts of truth: correspondence and coherence. But this should not prove too difficult an obstacle. There are serious reasons for using both these notions. Correspondence has intuitive plausibility and coherence seems needed to avoid scepticism. But coherence does not exclude the malin génie, there is always the possibility of falsity, but it is not reasonable to expect to be able to exclude its possibility. It is only if we take knowledge to require complete certainty that the exclusion of the malin génie would be required. But knowledge does not require

---

13. Walker, R.C.S.: The Coherence Theory of Truth, London, Routledge, 1989, pp. 6-7 and 210-220. Walker discusses theories of truth but essentially the same questions arise for a coherence theory of the justification of knowledge.

certainty or perfect assurance.[14] So even though it is correct that these two notions are at work in accounting for our knowledge, there seem to be good reasons for it.

What I have presented here is an internalist theory of justification, not a theory of knowledge. As mentioned at the beginning of the chapter, knowledge requires truth. We may reach truth through some natural relation between the knower and the known. But justification does not require truth. On the contrary, it can only supply us with evidence for truth, not with truth itself. This is not because justification is entirely divorced from any relation to truth. Justification is evidence for truth but the justification must be accessible to the knower. Beliefs are not only justified when true. The justification of beliefs is a function of the relation between a particular belief and other beliefs, primary and secondary. Coherence of beliefs is evidence for their truth. Coherentism explains how we are warranted in claiming to know.

This seems to me to follow from what has been said about knowledge and truth and that justification does not guarantee truth. If we add to that the claim that justification is uniform in the sense that we can have the same kind of justification for both true and false beliefs, then the account I have given of the relation between knowledge, belief, justification and truth becomes at least possible and, hopefully, probable.

---

14. Karlsson, M.: "Epistemic Leaks and Epistemic Meltdowns" in Hume Studies, Vol. XIV, Nr. 2, November 1990, p.122

3.2. Moral Knowledge

I have given an account of the justification of knowledge in general which does not refute the sceptic and hence it does not satisfy the strictest conditions we can lay down for the possibility of knowledge. The question is whether the notion of moral knowledge is more susceptible to sceptical arguments than the general notion of knowledge.

Moral knowledge is knowledge of a moral property. 'A knows that p' expresses moral knowledge when p is filled by propositions like 'x is wicked' or others, referring to moral properties. Now, one question which needs to be addressed is about the relation between natural or social properties and moral ones. Earlier I argued that moral properties were supervenient upon natural properties. What this relation comes to is that two things cannot differ in their supervening properties unless they differ in the properties supervened upon. This relation of supervenience can be construed in different ways, which does not matter in this context. But the important thing about this is that supervenience, as described, blocks one claim to moral knowledge. One might want to claim that if you knew all the natural properties of a particular situation, then you would know all you could possibly know about its moral properties. This is not true. The reason is that you would not know the moral properties. Only if you antecedently knew that in this

type of situation, this type of moral qualities occurred and there was no obvious defeating quality, could you say you knew the moral quality. We could also claim that you would have to know that this particular moral quality occurred in situations with these particular natural qualities. The only thing you could know on the basis of supervenience is that in two identical sets of natural qualities there will be the same moral quality. You could not know what the moral quality was.

Even if one accepted the account given here of the justification of knowledge in general, one might want to reject the possibility of moral knowledge. A number of arguments could be used to support this point of view. I shall consider four such arguments and attempt to reject them. In the course of discussing these four arguments I hope to make it plausible that moral knowledge is not to be considered impossible any more than knowledge in general is to be considered impossible.

The first argument is the argument from queerness as Mackie calls it.[15] This argument can come in two varieties. It can be ontological and it can be epistemological. It is only the epistemological variety that I shall consider here. The argument goes like this. If there were objective values or moral qualities, they would be entities of a very peculiar sort, unlike anything else in the universe. They would require a special faculty in us to discern them, different from

---

15. J. Mackie: Ethics, Harmondsworth, Penguin Books, 1977, p. 38-42

other ordinary ways of knowing. There is no evidence that we possess such a faculty. Hence we cannot know moral qualities.

This argument boils down to the claim that moral objectivism must assume intuitionism about moral knowledge and moral intuitionism seems to be implausible. The first question is why is intuitionism so implausible? I guess the response should be, according to the description above of the argument, that the claim about a special faculty is implausible. But why is that implausible? The answer to this comes in two parts. First, empirical knowledge comes through the senses and hence it is plausible to think that all our knowledge is related to perceptual faculties in a similar way. Second, there is no evidence at all for thinking that there might be an organ through which we could receive our moral perceptions and acquire our moral beliefs as we acquire our perceptual beliefs through the perceptual organs.

I take it to be obviously true that there is no organ detecting the presence of moral properties in the same way as our eyes, for example, detect visual properties. But a moral faculty is not the same as a moral organ. A moral faculty could consist of a distinctive way of seeing the world and a set of distinctive moral concepts. It does not require a particular organ. There is plenty of evidence for a moral faculty in this sense, even though there is no evidence for a moral organ.

One idea pervades the discussion of this subject and the analogy between perceptual beliefs and moral beliefs. It is that perceptual beliefs are probably the best candidates for foundational beliefs in our claims to empirical knowledge. It is often claimed that moral intuitions have this structural feature. But I have already argued that foundationalism in general does not give us the best theory of knowledge in general and there is no reason to believe that it gives us a good theory of moral knowledge in particular.

There is one argument against coherentism of the type here espoused. D. O. Brink argues for an internalist coherentism about the justification of knowledge. The notion of considered moral views plays an important role in such a theory. Considered moral beliefs are formed under conditions of cognitive reliability and also on the basis of impartial and imaginative considerations of the interests of the relevant parties.[16] This type of moral belief is used in justifying other beliefs and actions. One property of the set of considered moral beliefs is coherence and these beliefs are reliable and hence they can be evidence of moral truth. The end point of coherentism is reflective equilibrium and it is its end point because it is only reached after systematic justification. It is this end point which is contentious.

It seems to me, as I have said earlier, that the goal of systematic justification is not a reasonable goal for rational human beings because they are finite and

16. D.O. Brink: op. cit. p. 132

moral beliefs with all other beliefs are infinite. There is another reason for thinking that this is an unreasonable goal for moral knowledge. Knowing a set of facts p-px seems to imply that on a systematic justification all other facts or very many others, at least, must enter the justificatory chain and it becomes problematic where one's knowledge stops. Do we know, for example, all the propositions implied by our knowledge of the facts p-px? Even if we accepted a positive answer this does not mean that we would know all other facts. Even though the justificatory chain was deductive in nature, this would not follow. But the point is only that this blurs the borders of one's knowledge. So reflective equilibrium interpreted in this way is not a reasonable ideal. Our body of coherent beliefs is always a half way house, never fully systematically justified.

But it might be objected to using the notion of considered moral beliefs in this way that these beliefs are foundational, even though not claimed to be so. But to be foundational they would have to be self-justifying.[17] And whatever considered moral beliefs are they are not self-justifying. Even though there is no place for the notion of reflective equilibrium in the sense described in a theory of moral knowledge, there is no need to deny that considered moral beliefs forming a coherent set have a role to play. Such beliefs are credible, we have a reason to accept them as true, and they can be used to predict moral behaviour and they are

17. D.O. Brink: op. cit. p. 134

revisable. If we have a statement about some moral behaviour and it contradicts a belief which we have taken to be well established and a considered moral belief, then we must either reject the statement or revise the moral belief. In that respect considered moral beliefs can have a similar position to theoretical beliefs in science, they generate hypotheses and predictions which can be checked against experience. An example of a considered moral belief is that causing avoidable pain is wrong.

The second argument against my view is the argument from disagreement or relativity.[18] It goes something like this: there are radical differences in moral opinions and moral judgements. Hence, our moral judgements are very unlikely to be instances of moral knowledge. The idea in this, it seems, is that the principle of non-contradiction is violated, if we allow that it is possible for two claims to knowledge to be valid involving contradictory propositions. This is a reasonable thought because knowledge involves truth and the law of the excluded middle excludes the possibility that two contradictory statements can both be true at the same time of the same things. So a variety of different views on moral properties seems to imply that knowledge of those properties is impossible because all the views cannot be true.

This seems to me to be correct, as far as it goes. But the obvious question about moral opinions is: why

18. J.L. Mackie: op. cit. p. 36-38

should we take all moral views at face value? Why should we believe them to be true without further argument? These questions are relevant because it is only if all moral opinions are taken to be true without further deliberation that this argument is strong. One move at this point might be to say that the argument from disagreement only gets a grip when we add that the best explanation of our moral attitudes is that we approve of our practices because we partake in them not the other way around. This seems to imply that we believe murder to be wrong because we partake in social practices accepting this view. But this way of justifying the belief gets the chain of justification the wrong way around. We partake in such practices because we believe murder to be wrong. It is possible to evaluate such a belief independently of the practices.

The important question in this context is how do we know that any particular moral proposition is true? Is there any procedure we can rely on to distinguish between true and false moral beliefs? Implied in what has been said already has been the view that our ordinary moral beliefs have certain initial plausibility and these beliefs should be checked against other moral beliefs, secondary non-moral beliefs and moral principles when thinking about moral matters. If these beliefs can be shown to be coherent with all these other beliefs we have the best ground possible to take them to be true. This checking against other beliefs assumes that our initial moral beliefs play an important role in constructing a

moral theory, the relation between beliefs is inferential and the theory-construction process should result in a coherent end-state. If we find that the moral belief we are thinking about does not cohere with other beliefs of ours then we must either reject it or revise it in some way or revise other beliefs incoherent with it.

Is there any other procedure or method which could be used in thinking about moral decisions or to construct a moral theory? I shall mention two other suggestions. One is that our ordinary moral views should play no part in constructing a moral theory.[19] There can be various arguments for this view. One is that our ordinary moral views should and can play no part in constructing a moral theory or coming to a moral decision. One argument for this could be that some of our ordinary moral notions have a disreputable epistemological origin. Some of the moral beliefs we hold strongly we hold, not because of a justification or deduction from moral laws or principles, but because situations strike us in a particular way. Hence we need to start from well established first principles. Another method could be that our ordinary moral notions ought to play a larger role than on the coherentist view in the sense that not all our ordinary moral views should be revisable. They might act as constraints on moral theory construction.[20]

---

19. See, for example, R.M. Hare: Moral Thinking, Oxford, Oxford University Press, 1981, ch. 3 and 12
20. These views are discussed in W.D. Solomon: "Moral Realism and Moral Knowledge" in Proceedings of The American Catholic Philosophical Association, Vol. 59, 1985, esp. p. 46-49

These two alternative views deserve a separate discussion which I shall not go into. I shall mention two reasons to doubt the former view. The first reason can be seen from this question: why is it a reasonable requirement that no moral attitude should play any role in constructing a moral theory? This means, for example that Socrates should have dropped all his moral opinions in deciding whether to escape from prison and subject himself to some general method which might result in it being rational that he ought to have tried to escape from prison which he did not by any means want to do whatever his friends argued. This is only a reasonable requirement if all our opinions are on the same level, so to speak, that they are all preferences and no more than that. This is counterintuitive. The second reason is that this view implies that there is a one way traffic in revisability. Moral opinions can be revised in the light of general principles, not vice versa. But we can and sometimes do revise our principles in the light of specific examples. These two reasons will have to suffice as suggestions that this is not a promising method to construct a moral theory which explains moral knowledge.

The latter view is that some of our ordinary moral beliefs are such that they act as restraints on the construction of moral theories. This means that they are not revisable. Candidates for such beliefs could easily be 'torturing is abhorrent','murder is wrong' or 'rape is foul'. On the basis of these beliefs we could have corresponding absolute imperatives. It seems that rape

could not but be foul and murder could never be justified. If there are any absolutes, there does not seem to be any room for revisability of moral beliefs. So, if this is correct, coherentism, as here described, is more limited than I have supposed.

The difference between this method and coherentism is the priority given to coherence and revisability on the one hand and the priority given to some absolute, non-revisable beliefs on the other. They both agree on accepting our ordinary moral beliefs as having some evidential force. On coherentism's view the aim is to find a theory which is coherent and non-contradictory on the assumption that this will be the true theory and the justified one. One problem is that moral beliefs could conflict in the sense that they would entail two imperatives both of which we could not act on in a particular situation. Then it would be true that I ought to do A, I ought to do B but I cannot do both. One inference we might want to draw from this possibility is that a fully coherent theory of moral knowledge is not possible. Hence, we should allow some of our ordinary views to be constraints on theory.

This challenge to coherentism is not as serious as it seems. The first thing to notice is that these two views have many things in common. They accept the evidential force of our ordinary moral beliefs and they accept that coherence is a worthwhile aim, even though the absolute view does not give it unlimited priority. But the sticking point is the possible problem of

incoherence or inconsistency and this problem seems to undermine both realism and coherentism. I shall consider four responses to this problem. The first is that realist coherentism can accept that there are situations in which there are no correct answers to the question what is the right thing to do. But it points out that this does not imply that any answer goes. It does matter what we do in these situations. The second is that the scope of the principle of coherence is not all possible situations but most situations. It is precisely situations like moral dilemmas which limit its scope. Accepting moral dilemmas as possible does not mean that all moral situations are like moral dilemmas, only a limited number. The third response is to point out that it is perfectly possible to have a coherent non-contradictory set of considered moral beliefs and yet to have to decide between two options which are inconsistent in the sense that I cannot do both. The important point is that the inconsistency in question is not identical with being logically contradictory. The impossibility in question is an empirical impossibility. It is not the case that the descriptions of the two actions contradict each other or that they are self-contradictory. This empirical impossibility is in no way caused by or derivable from a contradiction in the set of considered moral beliefs. It becomes problematic when we add the admittedly plausible principle that ought implies can. Fourth, moral conflict is only problematic, if we think that moral realism implies that there is an algorithm or a computational

method to decide what is right in every possible situation. Moral realism implies no such thing.

The third argument against the view I have advocated is that moral beliefs do not allow prediction and confirmation in the same way as beliefs about natural properties do. The idea seems to be that in the case of non-moral beliefs we can better check whether the belief is true than in the moral case. The obvious case to compare is science. In science we have theories generating predictions leading to observation sentences and these observation sentences can be checked by experiments. Now the failure of an observation sentence to get confirmed does not necessarily lead to the rejection of the theory. In deriving the sentence from the theory we may have made a mistake in the auxiliary hypotheses used, for example. We check observational beliefs and they are not wholly independent of other beliefs. Scientific theories do not come isolated from observational sentences.

Why should it be thought that moral beliefs can not generate predictions and observation sentences? If we look at the matter, there does not seem to be any reason against it. If we know that A is a good man and he has promised to do x, then we can predict that A will fulfill his promise, even though it will be at some cost to himself. If A keeps his promise and no alternative explanations of his action are plausible, then I can conclude that his action tends to confirm the principle about promise keeping.

So observational beliefs have a role to play in ethical theories along with considered moral beliefs. This comparison, at least, does not show any break between ethics and science.

The fourth objection to the coherentism is that it leaves no independent criterion of the validity of moral theories or moral knowledge. This should lead us to reject moral realism. One argument for this could be the claim that the truth of a moral claim consists in its coherence with other beliefs. I do not claim that coherence is the only constituent of moral truth. I have already admitted that we must allow for correspondence in 3.1. Both these concepts play their part. Even though we had the best justification in terms of coherence there is no guarantee that we might not be mistaken. Systematic error is possible. Coherence alone can never guarantee truth because it can always lead to false beliefs. So the view advocated here does not identify moral truth solely with coherence, but it recognizes our limits in the sense that our best justification for moral knowledge may be coherence but it is not necessarily the whole story. There may be other possible arguments for this objection but I shall not discuss them.

3.3. Williams' Theory of Moral Knowledge

Bernard Williams has put forward a theory of moral knowledge which, if true, would refute the view I have argued for here.[21] It is worth looking at in detail.

Williams distinguishes between the scientific and the ethical and he argues that the difference between these two is best explained by convergence. In the scientific sphere we can ideally expect answers to a particular question to converge and this convergence should be explained by these answers representing things as they actually are. It is not impossible that a convergence on answers to ethical questions could take place but this convergence should not be explained in the same way as in the scientific field.

This distinction is related to another distinction. We can distinguish between the absolute conception of the world and the perspectival conception. The ethical is a subset of the perspectival. The absolute is not our notions of the world prior to any system of beliefs about it. We form our conception of the absolute on the basis of some of our beliefs about the world. But in forming this conception we concentrate not on the world out there but on how our beliefs about the world represent what they are about. In this way we form an absolute conception which is such that any inquirer would arrive

---

21. B. Williams: Ethics and The Limits of Philosophy, London, Fontana, 1985, ch. 8

at it, even if he was very different from us. The perspectival is constituted by properties which do not have this quality, properties like green. The claim is that the absolute conception would give a good explanation of how knowledge is possible of the absolute and the perspectival. The crucial idea here is explanation and in the perspectival field it is the different explanations of two or more perspectival beliefs which justify the distinction between different sorts of perspectival beliefs, like ethical beliefs and perceptual beliefs. In the case of perceptual beliefs a theory of error is possible which is not possible in the case of ethical beliefs.[22]

The question is how to account for ethical knowledge within this framework. It might seem fairly straightforward to deny the possibility of moral knowledge, but Williams' theory is much more sophisticated than that. To arrive at what he wants to say about moral knowledge we shall assume with him that we are dealing with a society that is "maximally homogeneous and minimally given to general reflection." This is the hypertraditional society in which its members simply use certain ethical concepts.[23] Can the members of such a society be said to have ethical knowledge? Williams does answer yes, but his answer is different from the one I tried to argue for earlier.

---

22. B. Williams: op. cit. p. 150
23. B. Williams: op. cit. p. 142

Williams distinguishes between reflective and unreflective knowledge and the distinction is roughly that reflective knowledge involves being justified in the sense of believing that one knows. In the hypertraditional society only the unreflective kind is available. Unreflective knowledge in such a society is problematic. One reason is that it is not easy for an outsider to determine the truth of the judgements in question because he cannot use those same concepts. But even though an outsider might be barred from using the ethical concepts in question, it is always available to him to say 'what they call F' and recognize what they say as true.

Parallel with the distinction between reflective and unreflective knowledge is the distinction between thick and thin ethical concepts. Examples of thick ethical concepts are concepts like 'courage', 'generosity', 'brutality'. Examples of the other kind are 'good', 'right'. The thin concepts can only be used at the reflective level but the thick concepts can be used at the unreflective level.

But the basic problem is how should we think of the relation between these two kinds of knowledge. Williams suggests two models, an objectivist model and a nonobjectivist model. On the objectivist model we see all judgements, both on the unreflective and the reflective level, as having general implications which the persons uttering the judgements may or may not realize. A member of a primitive society might believe that there are holes

in the sky letting in light or whatever. On the objectivist model we would want to say that this man's belief about the stars is false. On the nonobjectivist model one would look at the judgement as a part of a way of life, a cultural artifact, and there would be no objective grounding possible for ethical concepts. The argument for this seems to be twofold. On the one hand the objective grounding would have to be in terms of the thin ethical concepts and general arguments. But the thin ethical concepts are precisely not world-guided as would be necessary if they are to be accepted as objective. On the other hand, and as a result of this first point, the other available option for objective grounding is to use concepts from the surrounding culture. But these do not give an objective grounding in the required sense because they only serve you for a particular culture. There are no culturally independent, thick, ethical concepts.

This does not exclude the possibility of convergence in the ethical. The thick concepts a culture shares are world-guided in the sense that we would expect those sharing the same culture to share the same thick ethical concepts and know how to apply them. This is the case with colour concepts as well. But the explanation of the convergence of colour concepts is that we know that the perceptual capacities underlying our perception of the world present it to us in reliable and useful ways. This is not possible for ethical theory:

> "But while it (ethical theory) might explain why it was reasonable for people to have these various ethical beliefs, it would not be the sort of theory that could explain why they did or did not have them. It could not do something that explanations of perception can do, which is to generate an adequate theory of error and to account generally for the tendency of people to have what, according to its principles, are wrong beliefs."[24]

This is ultimately what justifies the distinction between the scientific and the ethical.

As should be clear Williams' view is a powerful argument, if true, against the views I have stated and attempted to justify in this chapter. The important things about it are the distinction between the ethical and the scientific and the rejection of the objectivist model which has been implicit in everything I have wanted to say up till now in this chapter.

The possibility of a theory of error was the important point about the scientific whether it was of the absolute or of the perspectival variety. The claim is that it is not possible to have such a theory for the ethical. Williams claimed that the absolute conception should be concerned with how our beliefs come to represent the things they do and in that way explain our knowledge. The idea seems to be that evolutionary theory,

---

24. B. Williams: op. cit. p. 151

the physiology of perception and the physics of light, sound or whatever have supplied us with such an explanation for perceptual beliefs.

The first thing to notice about these branches of science is that they are not aimed at explaining our beliefs. Their explananda are things in the world whatever part of the world a particular branch of science is concerned with. Physics explains physical phenomena, evolutionary theory explains the development of biological beings. If we look to cognitive science to explain our beliefs, the important thing about cognitive science is that it is independent of any particular theory in physics, for example. The results of cognitive science would obtain whether we believed in Aristotelian, Newtonian or the modern physics. So no results in physics can be adduced to confirm or reject a theory in cognitive science.

The theories in physics are confirmed or refuted by experiments, as in many other natural sciences. In some sciences we do not have this powerful tool for checking theories, sciences like meteorology. The important thing about this is that it is not the import of science that explains our beliefs, not even the beliefs of the scientists themselves.[25] The experiments can confirm or reject the truth of a particular hypothesis about the world, but where we do not have experiments there is

---

25. W. Quinn: "Reflection and the Loss of Moral Knowledge: Williams on Objectivity" in Philosophy and Public Affairs, Vol. 16, Nr. 2, p. 200-201

ultimately no further check possible than the ingenuity of the scientists at interpreting the data.

At this point we can distinguish between two senses of belief. One is a psychological sense and one a logical sense or scientific. This is the difference between belief as a mental event and the content of belief. I take it to be obvious that the truth of a content of belief cannot be derived from the occurrence of a mental event, unless, of course, the content of that belief were about that occurrence. The truth of the statement 'snow is white' is not determined by my belief that snow is white. It is not even determined by the fact, if it be a fact, that this belief is widespread. There seem to be three things going on here: it is the mental event, the content and the truth of the belief. As I have described this the question of truth can only arise for the content.

On Williams' theory the absolute conception should explain how it itself is possible and how we come to know it, it should explain why we do or do not have particular beliefs. The truth of those beliefs is supposed to play an important part in such an explanation. In the light of the distinction above it is not at all clear how beliefs as mental events are supposed to be connected to the truth of their contents. One explanation for saying above that the results of cognitive science are independent of results in physics, for example, is that cognitive science deals with the relation between the mental events and their content but physics is concerned with the

content of some beliefs and its truth or falsity. There is no necessary connection between the content of a belief and its occurrence as a mental event. Rational beings do have beliefs which they have acquired in a variety of ways. Some they get from their parents, some from their peers, some from their own thinking, some from books and some from other places like the media. Some of these beliefs are reasonable and some of them are, maybe, true. The way to find out is to see if they are justified.

The question now is: how is the theory of error possible? It is supposed to be embedded within a general theory which does not only explain that some beliefs are mistaken but is also supposed to explain how we come to acquire the beliefs we have, mistaken or otherwise. There is no problem in explaining how beliefs are mistaken; the mistaken ones are the false ones. But the general theory should do more. It should explain how we come to have these beliefs. The primary example of such an explanation is, I think, perceptual, especially colour, beliefs. The explanation would be similar to this: we have a theory of light explaining, for example, how light is different at different wavelengths. We have a theory of the nervous system which describes how the eyes and the brain link with light from the world around us. Given the natural functioning of the nervous system, particular wavelengths of light should be perceived as certain colours. The idea must be that we could in principle explain beliefs in the sense of mental events and we could also explain

systematic mistakes by, for example, deficiencies in the sensory system. This seems to me the case for Williams' theory of error for scientific knowledge.

There are some questions which should be raised about this account. First, suppose this is correct about colour concepts. Does it work for other kinds of concepts? Are colour concepts not special? I am not going to answer these questions, but they do point to the fact that further argument is needed. Second, I think there are reasons to believe that this account for colour concepts is not as convincing as it might appear. It is a fact that there is good consensus about the use of colour concepts. It is this consensus which makes the account above of the explanation of our beliefs plausible, not the other way around. Were there more divergence in the use of colour concepts than there is now, the plausibility of this account suffers.

If we can, and sometimes do, make mistakes in our judgements of colour, how would an explanation of such a mistaken belief look? It seems that it could not go beyond pointing out that the belief is false and could not give any account of why one came to hold such a belief. In the case of colour blindness it is possible to go further and give an explanation of why one comes to hold systematically mistaken beliefs about some colours. But the case of colour blindness is special and not typical.

Third, it should be noticed that a scientist inquiring into physiological events in our brain is

attempting to find out the type of events occurring and the relations that hold between them. There is no requirement for him to hold that such events enter into explanations of our beliefs. Nor would the findings of such a scientist enter into explanations of his own beliefs. He would hold his scientific beliefs on the best evidence available to him. Even though it is reasonable to expect to be able to reject or confirm our beliefs in the light of evidence, it is implausible to think that we can have an account of how we came to hold a belief in all cases. It also seems to be true that such an account of origin would not explain the truth or validity of the belief. This seems to me to essentially the same as the chimera of trying to explain the achievements in the arts and sciences by the childhood or the upbringing of the artist or scientist. This way of explaining perceptual beliefs, and other beliefs as well it seems, is the only good explanation to be had of our beliefs. The theory of coherence I have argued for in this chapter endorses that view. This seems to me to throw serious doubt on Williams' views on the distinction between the scientific and the ethical.

The other important thing about Williams' theory is his rejection of objectivism. The notion of objectivism in this context is the view that knowledge at a certain level, here the unreflective level, is impossible without consequences at another level, here the reflective level. The distinction between these two levels can be analyzed in terms of thick and thin moral concepts. The thick ones

can be deployed at the unreflective level but the thin ones at the reflective level. The thick ethical concepts are world guided, objective, but the thin ones are not. Reflection, on Williams' view, seems to consist in, at least in part, the use of thin ethical concepts. In starting to use thin concepts to reflect on a system of ethics, certain implications come to hold which did not hold before the reflection took place.

The key question, it seems to me, about this is why should we believe that the thin concepts are not as objective as the thick ones? The distinction between thick and thin ethical concepts is not very clear.[26] Sometimes it is a matter of generality and sometimes a matter of level. If it is a matter of level, the concept of goodness, which is at the second level, should not combine with concepts of the first level. We should not be able meaningfully to talk about a good parent or a good friend which are thick concepts. But the concept of goodness does not operate at a different level from these thick concepts and we have no problem in understanding the notion of a good parent, for example. If the thin concepts are more general than the thick concepts, why should that count against their objectivity? It is certainly true that the concept of goodness is applicable to a bewildering variety of things, but this does not mean that its meaning is unclear. The function f(x) does not have value until we have given a value to x. This does not mean that the function is unclear or in any way

26. W. Quinn: op. cit. p. 202-203

less objective than it is after x has received a value. The same seems to apply to goodness.

It is true that probably there is wider disagreement about thin ethical concepts than about thick ones. But as has been pointed out before, disagreement needs not to be a pointer to nonobjectivity or subjectivity. But if agreement were reached about thin ethical concepts, would the explanation of that agreement be in any way different from explanation of agreement about thick ethical concepts? Thick ethical concepts are world guided, objective. Presumably, something about the objects the concept is true of, be it their use or one or more of their qualities, should explain the convergence of beliefs and the agreement. The same should apply to the thin concepts. Admittedly, the concept of goodness may be difficult, if we tackle it on its own, but as soon as we take into consideration that 'good x' is like a function this should not be problematic.

Williams rejects any possibility of a theory of error for all ethical concepts. But it might be worthwhile to ask whether an error theory for thin ethical concepts would be different from an error theory of thick ethical concepts. A typical mistake in finding out whether an action is right or a man good would be to think something relevant which was not. If what I have said is correct, we should not expect any difference in the case of thin concepts. The coherence with other beliefs would be evidence of truth.

I do not claim to have refuted Williams' theory. But I hope to have shown that it is reasonable to doubt it.

3.4. Conclusion

In this chapter I have argued that a coherence theory of knowledge best explains our knowing the external world. Even though moral knowledge is more complicated the same considerations apply as for knowledge in general. Williams' theory does not show that this account of moral knowledge is false.

Chapter 4

Moral Expertise

I have just argued that it is reasonable to believe moral knowledge to be possible and justifiable. One natural inference from this is to take moral expertise to be possible as well: where knowledge is possible expertise is possible. In this chapter I shall argue that moral expertise is indeed possible and that there can be moral experts. But to make this thought plausible much more will need to be said about the nature of moral expertise and why so many people find it repugnant.

4.1 Moral Expertise and Wisdom

To be expert at something is to be knowledgeable about it, to be skilled at it or to have had wide experience of it. Expertise is expert opinion or knowledge. This explanation of 'expert' and 'expertise' can be found in any dictionary. I checked The Shorter Oxford English Dictionary. What is obvious from this is that there is a direct connection between knowledge and expertise, which should not be surprising. It also seems that expertise is closer to skill than to the cognitive aspect of knowledge. This should not be taken too strictly because it seems fairly clear that some experts have no other skills than cognitive skills as parts of

their expertise and some experts have other skills as a part of their expertise. An economist must understand some part of the workings of the economy well and be able to explain these workings if he is correctly to be called an expert. This is as pure an instance of knowledge as we can get. A cook is an expert at preparing food for eating. This does not mean that he will be able to explain why a certain mixture of ingredients turns out be delicious food, only that he knows how to mix them, that certain things go well together and others do not. In short, he knows what makes good food. Cooking is a skill and knowing how the economy functions is not a skill in the ordinary sense of the word. But I do not think that one should hesitate to call a knowledgeable economist an expert nor to call a cook an expert.

Sometimes the use of the term 'expert' in ordinary language is so loose as to mean anybody who knows anything about something. But the more established idea of an expert is someone who is knowledgeable about something or about some particular area of knowledge. This description would include the cook and the economist. We would expect an expert to know the general principles of his area of expertise, how to apply them and at least some of the nuances of such application. He should know all the major theories held and the viewpoints on the serious issues in his field. In knowing the theories he knows the assumptions made and the inferences drawn from these assumptions. The expertise in question can be learnt and this learning results in a

change of ability. This description of expertise seems applicable to any field of knowledge whether you are talking about an economist, an historian, a linguist, a doctor or an engineer. One might hesitate about the cook, as I described him above, and ask whether his skill required any theoretical knowledge. It is certainly true that some of the things do not apply to him, but this should not exclude him from the field of experts because being a good cook is dependent on skills. Included in his skills is knowledge of the ingredients he uses, even though he might not be able to explain why these ingredients go well together. I do not think it would be possible to have a good cook who knew nothing about the properties of the ingredients he uses. Knowing what an onion is just is knowing some of the properties of an onion. A cook knows those properties of an onion relevant to cooking.

One part of the idea of an expert is that the expert must know more than someone else, a layman. In a field of knowledge it is possible that all members of a group have the same level of knowledge. It is certainly possible that all members of a society have the same skills in cooking. Then it has no point to distinguish between experts and laymen. But we can say whether the level of knowledge is high or low, whether the members of that society should all be called experts or not. So it seems that expertise is relative to the distribution of knowledge within a group and also relative to the level of knowledge acquired by a group. If we had a group of

people and the large majority did not know how to boil potatoes, let alone anything more complex, then anyone who knew how to prepare ratatouille or to roast beef would be an expert.[1] But if the majority or most of the group knew how to prepare ratatouille and to roast beef, then those who knew how to cook exquisite, complex dishes in addition to those practically everybody else knew, would be the experts. The relation between the experts and laymen in this latter instance is not identical to the one in the former. The level of knowledge makes a difference.

Moral expertise is being knowledgeable about moral properties. The description above of expertise seems to apply to moral expertise. A moral expert is somebody who is acquainted with moral theories, the arguments for and against them, he should be able to discern what are reasonable assumptions about actions and persons and the inferences that can be drawn from them. He should be able to think rationally about moral issues and understand the nature of moral concepts. He should be able to use analogies and be able to tell what are the relevant differences and similarities between two things being considered. He should be able to recognize the nuances in applying the general principles of the moral theories, seeing what is justified and what is not. He should be

---

1. The disparaging remarks about cookery Plato makes in Gorgias, 463a-465d, are quite unjustified, it seems to me. There is no obvious justification for saying that cookery is unreasoning and hence does not have a rational account. The question is about the nature of the rational account to be given of cookery.

well acquainted with the major issues about morality and the viewpoints on these issues and how these are related to the theories. Moral expertise also requires experience in dealing with people in various situations and sensitivity to their needs and feelings.

One thing that might occur to you about moral expertise so described is that it must be closely related to wisdom. But what is wisdom? The Shorter Oxford English Dictionary gives this meaning: "Capacity of judging rightly in matters relating to life and conduct; soundness of judgement in the choice of means and ends; sometimes, less strictly, sound sense, esp. in practical affairs." It is quite clear that 'moral expertise' is not synonymous with 'wisdom'. The obvious difference is the acquaintance with theories and arguments mentioned in the description of moral expertise which is not referred to in the definition of wisdom. But judging rightly in matters of life and conduct is certainly the hoped for result of achieving moral expertise. The objective of moral expertise must surely be that it is more likely that our moral judgement be right as a result of acquiring moral expertise.

Moral expertise, it seems, could be instrumental in bringing about wisdom. Wisdom consists of reflectiveness and judgement in practical matters, it is often proverbial and is sometimes implicit in what is said or decided rather than explicit.[2] By reflectiveness, which

2. See B. Blanshard: "Wisdom" in P. Edwards: The Encyclopedia of Philosophy, Macmillan, London, 1967, Vol. 8, p. 324-326.

did not appear in the definition above, is meant the opposite of impulsiveness. The idea is that conduct resulting from impulsiveness is more likely to be misguided than conduct based on considering the grounds and consequences of every action. This reflectiveness does not necessarily result from knowledge of moral theories or their application but from experience of rational beings in various situations in their lives, how they react and what they believe. It is obvious that such responses will have many things in common and be very variable. This indicates the difference between moral expertise and wisdom hinted at earlier. It seems that wisdom does not require any knowledge of moral theory but a sensitivity of mind and a willingness to think about one's own and others' moral reactions and beliefs.

Before going on I want to make one distinction which is important to avoid confusion. Sometimes moral experts are taken to be philosopher kings of the Platonian type, able to discern everybody's needs and real longings and willing to enforce their knowledge of other people's needs on them whatever their own beliefs and assessment. So if people did not accept the conclusions of the moral expert, he would be right in coercing them to accept these conclusions. A moral expert in the sense of a philosopher king is a paternalist and thinks that his paternalism is justified because of his superior knowledge. But a moral expert need not necessarily be a philosopher king. As I have described moral expertise, and as will become clearer in what follows, there was no

hint of paternalism being a part of it. A moral expert would certainly have superior knowledge to those who could not bother to think about moral matters seriously. But this does not entitle the experts to get others to accept their views by force. There are certainly connections between the possession of moral knowledge and being justified in attempting to change people's moral views. There are also connections between moral expertise and justifiable paternalistic intervention with someone else's action. If the intervention is non-coercive and based on moral expertise, there does not seem to be a presumption against it.

4.2 There Are No Moral Experts

Mary Warnock has argued that there are no moral experts. Implied in this claim, as presented by her, is that there is no moral expertise. It is instructive to see how she argues for this conclusion.[3]

Warnock uses four arguments, it seems to me. I shall describe them all first and then discuss them.

1. She states that moral experts, similarly described as above where I described moral expertise, are really no moral experts. They look like intelligent first-year undergraduates in philosophy. So this description does not capture anything like an expert or expertise.

---

3. M. Warnock: A Question of Life, Blackwell, Oxford, 1985, p. 95-100. Her views are endorsed by D. McNaughton: Moral Vision, Blackwell, Oxford, 1988, p. 203-205

2. Moral experts cannot be so called because the only point in calling them experts is that "their conclusions should be accepted without question."[4] The laymen must feel that they are in no position to disagree when the experts have delivered their judgement. But as this is not the case with moral experts there are no moral experts.

3. Everyone must make their own moral choices. Matters of family, life and death are at the heart of any morality. Decisions in these matters are based on everyone's conscience and conscience is the highest authority. Hence no moral experts.

4. In moral matters there can be better or worse judgements, but there cannot be a correct judgement. "That being so, no judgement can be imposed by one person on another as the only right or possible or proper judgement to make (even though each of us may feel that his judgement is manifestly best)."[5]

These arguments seem to lead to the conclusion that no moral expertise is possible and consequently that no moral experts are possible. Warnock does not clearly distinguish between these two claims. But it is possible to hold either one or both of these claims to be true or either one or both to be false. It is certainly conceivable to hold that moral expertise is possible and denying that moral experts are. Moral expertise might be so difficult to acquire, for instance, that it was not

---

4. M. Warnock: op. cit. p. 96
5. M. Warnock: op. cit. p.96-97

reasonable to expect anyone to achieve it. It seems to be contradictory to hold that moral experts are possible but not moral expertise. If no moral expertise is possible, it seems to follow that no moral experts are possible either. To have experts without expertise seems to be contradictory. Lastly, it is possible to deny both claims to moral expertise and moral experts. Mary Warnock seems to take the last option and, judging from the arguments I have described above, that is reasonable enough. It seems to me, though, that all these arguments are invalid.

The idea in the first argument is that the description given of the group of experts and the nature of their expertise does not capture anything resembling that. The capacity is that of first year undergraduates in philosophy, reasonably competent. But is this true? Do first year undergraduates feel confident in their understanding of moral concepts and the major moral theories? Are they also acquainted with the major moral issues and the viewpoints taken on them? This seems to be false.

But I do not think any heavy weather should be made of this point. Mrs. Warnock clearly does not intend it very seriously and it is closely linked to the second point.

The second point was that experts' conclusions should be accepted without question. If true, this is an important point. But first this needs some clarification. The claim is that the opinions and conclusions of the experts, apparently in virtue of their expertise, should

be accepted by laymen without question. Implicit in this is that there is no such acceptance of experts' view of moral problems. There seem to be two ways to doubt this. One is to point out that there are obvious cases of conclusions about moral problems where you would find such acceptance. The other is to draw attention to cases of expertise where laymen are as willing to question the verdict of other types of experts as they presumably are to question the verdict of moral experts.

Let us first look at verdicts or conclusions about surrogacy, which is one of the subjects tackled in the report of a committee Mrs. Warnock chaired inquiring into human fertilization and embryology.[6] Let us assume that after some thinking we come to the conclusion that surrogacy contracts through commercial agencies should be made illegal even though surrogacy arrangements by other means are not recommended to be made illegal. The arguments for this could be the welfare of the children born by a surrogate mother, the danger that their status would be very uncertain if disagreements arose, that a woman's labour should not be made a commodity like any other on the free market, nor should children be made a commodity. One could also claim that commercial surrogacy could seriously affect the family of the mother who undertook it.

These seem to be moral arguments for moral conclusions. Anyone who does not know a thing about surrogacy would not be in a position to dispute these

6. M. Warnock: op. cit. p.1-94

findings and would in all probability accept them without question. If the only point of calling someone an expert is that he can come to conclusions of this sort, then moral experts seem to be possible and actual, as these are actual conclusions. And they are not, apparently, hotly debated by laymen. It seems to me that a large part of the population would have no hesitation in accepting verdicts like these and not want to question them, maybe a majority.

Is it the case that in all undisputed cases of expertise the laymen are always willing to accept the verdicts of the experts without question? If a doctor advises his patient to stop consuming fats, as much as he can, because his heart is so weak, it is to be expected that the patient accepts the advice. But some do not and prefer to stick to their lifestyle rather than prolong their lives. There is nothing peculiar or strange in doing that. A meteorologist may predict rain tomorrow, but there is nothing untoward in disputing his prediction. Both these cases are cases of expertise in an unproblematic sense and there is no absurdity in disputing its verdict.

One reply to this could be to point out that Warnock said that the experts' conclusions "**should be** accepted without question." It is the force of this 'should' that ought to be considered. First of all, I suggest that 'without question' should be dropped without further ado. It simply is inappropriate as the cases above showed. It seems to me a condition far too strict and ruling out

practically anything we are usually ready to accept as expertise. Second, the obvious understanding of 'should' in this connection is that it is grounded in superior knowledge and therefore the verdict is to be accepted in virtue of it. If the verdict of experts is based on superior knowledge to our own, then we are to accept it. The problem with this interpretation is that it applies both to moral experts and to other kinds of experts because moral knowledge is possible, as argued in last chapter, and moral experts have, presumably, acquired it. This second argument of Warnock's does not support her view.

But Warnock's idea could be that those who accept the advice should not accept moral advice without question. This is compatible with believing in moral expertise and moral experts. The idea seems, then, to be that it is incumbent on those who receive moral advice to question it because they are the ultimate authorities in decisions about their own conduct. This is essentially the same point as the next one, to which I will turn. But the obvious reply to this is that it is true that everyone is the ultimate authority about what he decides to do, but this does not mean that he is the ultimate authority on whether his conduct is morally right or wrong.

Warnock's third argument was that matters of the family, life and death are at the heart of morality and everyone both wants to and must make up their own mind about them. It is also important that the Protestant

tradition in Britain, and in other North-European countries, is that the conscience of the individual is the highest authority in moral matters.

It is certainly true that no one can go through life without taking moral decisions about some moral matter or another. These are in most cases connected to the family. I am not so sure about life and death. In a sense one could say that one decides every day to go on living. But this is no real decision in practically all cases but just what you do every day. When one needs a real decision about whether to live or die, it can be excruciatingly painful. I do not think that people commit suicide lightly. There are other types of decisions about life and death. Some people may have to decide the fate of their parents or children, if they become very ill and come to the point of no recovery. I think it is untrue to say that people want to take such decisions. This is obviously untrue in the sense that everyone probably prefers to avoid getting into such situations, but it seems also to be untrue in the sense that not all would want to take such a decision if landed in circumstances of this type. In the present state of medical technology it seems unavoidable that some people will land in such situations. It is certainly possible that the medical staff took such decisions, but I believe that one rationale for putting the obligation on the closest relatives is that they are most likely to take decisions based on the interests of the patient. If it turned out that such decisions tended to disregard the patients'

interests, then this arrangement would and should be considered and changed. There is nothing inviolable about the person most concerned taking these decisions.

The principle that the individual's conscience is the supreme judge in moral matters is problematic. This is so because it ultimately amounts to everybody being the supreme judge in moral matters and no further appeal possible. If one accepts that propositions about morality can be true or false, then it is certainly possible that some of the judgements based on conscience could be false. This is especially clear if one considers that it varies among individuals what their consciences tell them to do. If moral judgements conflict, both of them cannot be true. This applies also, if one accepts that one moral judgement can be better than another. The inference one should draw from this is that the individual conscience cannot be the supreme arbiter in morality, in deciding what is the morally right action. It is the supreme authority on what each individual will do. The reason is that we can have true or false statements about moral rightness or wrongness. Decisions about actions cannot be true or false.

Another problem with this view of the role of conscience appears when we think about complex moral problems. Many of the moral problems confronting us in everyday life are relatively simple, and reasonable upbringing, basic education and some intelligence should equip us for solving them. This does not necessarily mean

that these problems are easy; they may be very difficult because of the importance of the issues involved.

Moral issues in human fertilization and embryology in the present time are becoming very complex.[7] Sexual intercourse, which can be considered relatively simple as a moral problem, is no longer the only way of begetting children. What should we think of these new ways of begetting children? To come to a conclusion about that is no simple matter and conscience is not a good or reliable guide to an informed and rational view. There are no established rules to rely on. What we need is to gather information, organize it, apply the moral concepts to these new techniques and draw some inferences about the morality or immorality of these new ways of begetting children. This can lead to us changing our original views on this matter or on other related matters, trying to make them consistent and, hopefully, right. The point is that conscience is not the best guide to intelligent views in a matter like this and it would be foolish to accept it as a supreme guide in complex moral matters concerning human fertilization and embryology. So the status of conscience as a supreme arbiter is problematic and it seems to be false to say about all moral issues that conscience is the supreme arbiter.

The fourth argument was that in moral matters there cannot be a correct judgement and so no moral judgement can be imposed by one person on another. This point is

---

7. see M. Warnock: op. cit. and P. Singer and D. Wells: The Reproductive Revolution, Oxford, Oxford University Press, 1984, p. 13-189

simply a confusion of moral expertise with paternalism. Whatever the connections between these two it is not the case that moral expertise just is paternalism. Experts of any kind are not entitled to force their advice on anybody simply in virtue of their expertise. But they can expect that their advice be taken seriously and accepted or rejected, as the case may be, on serious grounds. It is instructive to consider the report of the Warnock Committee. It is as clear a case as one can expect to find of moral advice to Government. This type of advice does not exclude statements of fact. The bulk of the report is a collection of statements about various important and relevant facts bout alleviation of infertility, about laws or the lack of them in this area and the possible future developments in this area and recommendations about the shape of the laws. The principles on which this advice is based are moral and can be seen, for example, in the advice that experiments on embryos are believed to be justified in the first two weeks of their lives.

Another important fact about this inquiry is that it is about an area which is new, as I mentioned above. So what is needed is an interpretation of the moral concepts we use to apply to this new area. In such a case it is not unreasonable that laymen should be willing to accept the verdicts of experts in this field and would find it difficult to disagree with them. Yet, Warnock expresses an important truth about the recommendation of such a committee: "It has to be remembered that Ministers are

not bound to follow advice they are given by committees of inquiry. The report is only advice. The recommendations begin to have force only if they seem sensible and persuasive to Ministers and their civil servants."[8] This a good description of the function of moral experts and avoids the confusion between moral expertise and paternalism. The important point to notice is that moral experts, like other experts, give advice. It is up to the one who gets the advice whether he accepts it or not. The acceptance depends, presumably, on the quality of the advice.

This concludes my discussion of Mrs. Warnock's argument against moral experts.

4.3 Further Objections to Moral Expertise

Mrs. Warnock's objections do not by any means exhaust the objections that have been raised to moral expertise. I want to start by taking a look at what might be construed as Mill's objections.

Mill's theory of individuality provides him with a justification of his opposition to paternalism and, apparently, to moral experts, if not to moral expertise. I will argue that this interpretation of Mill ought to be rejected.

The general ideas in Mill's theory of individuality seem straightforward enough even though he may be working

---

8. M. Warnock: op. cit. p. 99

with more than one notion of individuality.[9] He thinks that individuality is of the highest value and should be developed and encouraged as much as possible. Individuality consists in deciding and choosing one's actions and plans of life. In choosing one must deliberate about other alternatives to our plan of life and check their characteristics and come to a conclusion. This choice should assist our development rather than hinder it. Freedom to choose is essential, if the nature of everyone is to develop and for any choice to be meaningful. To achieve happiness one must develop one's character or individuality. Happiness on this view is not the satisfaction of a desire but the developing or perfecting of one's individuality.

Mill's theory does not presuppose the absence of objective truth in this area as might be believed. If it is up to the individual to choose his own plan of life, it might be inferred that there could be no objective truth about such choices. This would mean that any ideal an individual happened to embrace should be welcomed because it is his ideal. But this is not Mill's view. It is both more congenial to his general view to say that there is truth to be had in choosing a plan of life or deciding a moral action, but admittedly such truth depends on one's own characteristics and hence no general truth is to be had about what life plan anyone ought to

---

9. J.S. Mill: On Liberty, ch. 3 in Three Essays, Oxford, Oxford University Press, 1975, p. 69-91. See also C.L. Ten: Mill on Liberty, Oxford, Oxford University Press, 1980, ch. 5, p. 68-85.

choose. This is in accordance with his statements about polygamous marriages, for example, of which he disapproves strongly.[10] Such marriages should be tolerated but they are outside what can be considered attractive to human development. Another reason for thinking that there is truth to be had in these matters is that the choices one makes are in part discoveries and can be mistaken. In virtue of a false belief about my potentialities and unique endowments I can make a choice which can turn out to be mistaken.[11] So the choices one makes lead to self-knowledge and knowing oneself is a useful thing both to oneself and others. For these reasons it is false to say that Mill rejects the notion of truth in the personal sphere.

In choosing our own plans of life, in deciding our own actions in the light of our own dispositions our actions are our own. This is a key element in Mill's theory of individuality. But what does it mean to say that an action is mine in this context? It is clearly not sufficient that I performed it. It is necessary that I decided on it and it also seems to be necessary that such actions are "expressions of his own nature, as it has been developed and modified by his own culture."[12] From this I think one can infer that an action is one's own when one has freely accepted a personal ideal and tries to conform to it and this ideal is appropriate in the

---

10. J.S. Mill: op. cit. p. 112-113
11. See J. Gray: Mill on Liberty: A Defense, London, Routledge and Kegan Paul, 1983, p. 79
12. J.S. Mill: op. cit. p. 74

sense that it is based on an accurate perception of one's own talents and potentialities. This is admittedly rather vague but it is clear that on Mill's view an action is mine if it results from a chosen ideal and a settled disposition of mine.

Mill uses this to support claims about the badness of habit and how it prevents development. "The human faculties of perception, judgement, discriminative feeling, mental activity, and even moral preference, are exercised only in making a choice. He who does anything because it is the custom, makes no choice."[13] Implied in this is that a customary action excludes choice and hence cannot count as really mine in the sense described.[14]

Some of the implications for moral experts and moral expertise need to be drawn out from this account of Mill's theory of individuality. The first thing to notice is that Mill does not need to deny that moral expertise is possible. On this account of his theory it seems to me that moral expertise is perfectly possible. It may be that it will never amount to anything resembling a full blown academic discipline because any generalization will

---

13. J.S. Mill: op. cit. p. 72
14. It is arguable that this relies on too narrow a view of customary action and even a confusion of habitual and customary action. See Scruton, R.: "Freedom and Custom", in Griffiths, A.P.: Of Liberty, Cambridge, Cambridge University Press, 1983, pp. 192-195. It is fairly clear that if we give tradition and custom the place Scruton wants moral experts would have a larger role to play than I am willing to argue. This seems to be true even if allowing for the fact that deep desires and real interests must take into account the agent's perspective. But I will not pursue this argument but note that Mill seems to be willing to accept a substantial role for moral experts.

be difficult. But as a practical subject where the knowledge in question need not be backed by a full blown theory of any sorts, I do not see any reason for Mill to deny moral expertise. Second, Mill need not deny that moral experts are possible either. They would be the ones in possession of moral knowledge. The argument for saying this is that Mill allows that there are competent judges in distinguishing between higher and lower pleasures and that the higher ones are better for human beings.[15] This competence seems to be a competence in judging what is morally good. This is the sort of competence a moral expert would need.

It is possible that someone might want to reply on Mill's behalf that it is undesirable to seek the advice of the moral experts. The obvious reason is that it is the individual himself who should attempt to solve the moral problems he encounters. Otherwise he would not be fully rational and not attempting to develop his talents and faculties to the full. The defender of Mill might also want to say that seeking moral advice is unworthy of a rational being, somehow demeaning for the one who seeks it. He might even go further and say that seeking moral advice is morally crippling.

Before replying to these points I want to say that this is not a view that Mill seems willing to take. I stated this above in connection with those competent to discern the difference between higher and lower

---

15. J.S. Mill: Utilitarianism in M. Warnock (ed.): Utilitarianism, London, Fontana, 1962, p. 259-261.

pleasures. But it is clear from what Mill says about the doctrine of individuality that he is in no way opposed to moral guidance by men of genius or originality. What he is opposing with this doctrine is the tyranny of the masses and public opinion. He says:

> "The initiation of all wise or noble things, comes and must come from individuals; generally at first from some one individual. The honour and glory of the average man is that he is capable of following that initiative; that he can respond internally to wise and noble things, and be led to them with his eyes open. I am not countenancing the sort of 'hero-worship' which applauds the strong man of genius for forcibly seizing on the government of the world and making it do his bidding in spite of itself. All he can claim is, freedom to point out the way."[16]

Now back to the argument described above. It seems to me that this argument against the seeking of moral advice does not work. I shall look at the points in turn. The first was that somehow rationality requires that each individual himself tries to solve his moral problems. This is certainly true in the sense that each individual ought to take his own moral decisions if he wants to be a rational being, and as I have argued above, moral advice does not in any way encroach upon our ability to decide.

16. Mill, J.S.: On Liberty, op. cit. p. 82

As is likely, it will be the more complex moral problems that will prompt us to seek moral advice. Moral problems can easily be so difficult that we are at a loss what to do. If we do not seek advice, this can lead to our inaction. But our inaction may have worse consequences than our informed action. Rationality seems to require that we aim for better consequences of our actions rather than worse. Hence, rationality seems to require that we seek moral advice.

Let us accept then that there are such situations where we should seek moral advice. But it can be important how often such advice is sought. If it becomes a habit to look for the guidance of the moral expert, whenever a moral problem arises, it is likely that there comes a time when we would simply lose our ability to think and solve moral problems. And in so far as this ability is necessary for our moral development, we should try to use it as often as possible and in as various circumstances as possible. In some we would better seek moral advice. But doing that too often blunts our moral sensibility and our ability to solve moral problems. But it is a fine judgement where to draw the line but it seems reasonable that such a line could possibly be drawn and should be drawn. We should also remember a distinction between personal problems and social policy issues. There is no problem in arguing that on social policy issues, in so far as they involve moral issues, moral advice should be sought. It is in the personal problems where a line needs to be drawn.

The next issue was whether seeking moral advice was demeaning. But this can only be demeaning because it is presupposed that each individual should be able to solve their moral problems. But why should we think this? One view could be that everyone ought to be able to take his own moral decisions. In one sense this is a truism. Everyone can obviously take a decision in any matter, moral as well as others. But it is not true that everyone can take a rational decision in any moral matter confronting him without seeking information, deliberating, and seeking advice is part of acquiring information. If all moral problems were simple, like a decision what to have for dinner is simple, then it would be a reasonable requirement about moral decisions that everyone should be able to take them on his own without any advice. But if moral decisions are complex, like a decision to become a surrogate mother is complex, then rationality seems to require that moral advice be sought. So seeking moral advice cannot be demeaning because everyone should be able to solve their own moral problems.

To confirm this we can look at the seeking of advice in other fields. If you encounter a legal problem, let us say about the selling of a house, then you go to a lawyer and he will tell you what the law is concerning this issue, what the likelihood is of your achieving the desired objective and he might take into account the consequences of pursuing your preferred course. For those not trained in law this would be the only rational way to

conduct their case. Why should it be different in the moral case? I think it would be absurd to say that it is demeaning to seek moral advice on the moral aspects of a social issue. This only seems to be appropriate when talking about simple, personal, moral problems. One might have the feeling that seeking moral advice about simple, personal problems is a sign of failure on the part of the one who seeks the advice, just like going to a psychiatrist is sometimes considered to be a sign of failure. In a sense it can be a sign of failure because the problems may have grown too difficult for you to solve on your own. But then the best way is to seek help and I do not see how one could argue that it is in any way shameful to do so. Also, you do not need to be in any desperation or about to lose control over your own life to seek moral advice. You can be doing it before any real trouble arises because you think you need some advice. In seeking moral advice you are not failing, but rather doing the rational thing. Unless, of course, you were failing to take most of the simple moral decisions you have to take in everyday life. If it could reasonably be shown that seeking moral advice would probably lead to such a decrepit moral state, it should not be sought. This is really the argument that moral advice will cripple.

The question is whether it is likely that the seeking of moral advice will cripple you as a rational, moral being. This is only likely, if we are speaking about personal moral problems, not social issues with a

moral aspect. Also, it seems, that we can only talk about a crippling effect if the seeking of moral advice made you unable or severely diminished your ability to take rational decisions about simple, rather than complex moral problems, admitting that there is no sharp line to be drawn between simple and complex in this context. One of the decisions that most people have to take at some stage in their lives is about whether to marry or not. Nowadays, for many people, this decision is independent of the decision to live with someone. But let us assume, here, that there is only one decision. First, the marriage contract is not very clear at this point in time, even though, I guess, most people would expect that it covered sexual fidelity, mutual support, respect and love and doubtless various other things as well. Second, the decision to enter marriage is a moral one. It is a decision about how you will treat another person. Third, the decision to enter marriage is about a complex moral issue.

It seems to be true that it is preferable that everyone should ultimately be able to decide in his own case whom to marry rather than having to ask someone else to decide it for him. The reason is that it is he or she who has to live with the consequences of the decision. Other societies arrange the taking of this decision differently, usually by parents. There can be various different arguments for such an arrangement, the obvious one being the more mature judgement of parents which, we could argue, led to happier marriages. There might also

be a different arrangement suggested once by Samuel Johnson: "Marriages would in general be as happy, and often more so, if they were all made by the Lord Chancellor."[17] But this is beside the point here. The only thing I want to consider is whether it would be likely that the seeking of moral advice on a decision like this would cripple our moral sense and judgement. I guess many people seek advice on a decision as important as getting married. They discuss it with their parents or friends. Some of the advice they get will be moral advice because it involves how you treat another person and kinship ties and sexual relations are subject to considerations of justice. It seems that he who seeks such advice is possibly in a better position to appreciate what he is deciding to do than another who does not seek it. To that extent he is rational and has a sharper and deeper appreciation of the morality of such a decision. It is difficult to see how this can blunt our ability to take moral decisions. Admittedly, this is a decision about a complex moral issue rather than a simple one.

It seems to me that there is nothing quaint in seeking moral advice about social issues and complex, personal, moral problems. What about simple, personal, moral problems? I shall assume that avoiding causing unnecessary pain is a simple moral truth and that when we come across a gratuitous infliction of pain, as when a

17. As quoted by Michael Walzer in Spheres of Justice, New York, Basic Books, 1983, p. 227. The quotation is from The Life of Samuel Johnson by James Boswell.

cat is flayed alive, there is no question that it is wrong and that we should try to prevent it if there was any possibility of doing so without putting ourselves in serious danger. This seems a reasonable case of a simple moral problem. One of the reasons for calling it simple is that we do not need any advice on how to understand or respond to it, if it arises and if we are ordinary, normal, rational, human beings. When we reflect about this example, it is not at all simple to explain why we think the moral properties are as they are. But in this sense there is no simple moral property, any more than there is no simple physical property when closely scrutinized. But the main point about simple moral problems is that there are such things and we take it as a matter of course how we should respond to them. We do not need to ask anyone how we should.

If we do need advice on how to respond to simple moral problems, it is a sign of considerable weakness, which may go to the heart of your rationality and morality and make you unable to deal with ordinary problems of any kind. In this sense not being able to respond spontaneously to simple moral problems is indicative of a serious failure. It seems that being adult and fully rational requires that we can deal on our own with simple moral problems and moral problems that are likely to arise in any life such as moral problems within the family. We should not need to ask whether we ought to avoid killing other human beings intentionally.

But if the need is there, the rational option is to seek advice and help.

There are three more arguments against moral expertise which I want to consider in this part of the chapter. The first is that moral expertise seems radically different from other types of expertise. If true, there might be an argument for saying that there is no moral expertise. Knowing how to be brave or knowing what is morally good is not really knowing anything in particular. Knowing how to play golf means that you know how to hit a drive, how to putt etc. Knowing what is good does not entail any comparable identifiable skill. "We attribute technical goodness to a man on the ground that he is good at some activity. But there is no specific activity at which, say, the courageous man must be good - as the skilled chess-player must be good at playing chess and the skilled teacher must be good at teaching. There is no art of 'couraging', in which the brave man excels."[18] Von Wright takes this to imply that specific virtues do not imply specific activities nor act categories. Yet he does not deny that morality, the virtues, can be learnt.

There are ways of countering this. One answer is to say that the good man is "good at discerning what is right and wrong, and in doing something about it. He, for example, takes extra pains to find out if he hurt that old lady's feelings and to do something about it if he

---

18. G.H. von Wright: The Varieties of Goodness, London, Routledge and Kegan Paul, 1963, p. 139

did."[19] Being morally good or virtuous is difficult and needs to be learnt. We seem to have a natural tendency for the pleasant and the easy in preference to the arduous and painful. We need to learn that this tendency can be a source of moral failure. Moral failure sometimes results from wickedness, but more often from ignorance and weakness.

Another reply could be this. Von Wright's view of the nature of moral expertise ignores moral problems and moral dilemmas. It is only when moral principles clash that clear thought and experience in judging what is right and wrong become important. Even though no particular type of competence corresponds to every virtue, it does not follow that virtues do not require any skills. It is just a matter of judgement in every particular situation what competence and what skill is required.[20]

This requires some consideration. What is von Wright denying when he says that bravery does not need 'couraging'? One way of understanding this remark is to say that bravery does not entail any particular physical movement of the body. The problem with this is that it seems to be a truism which nobody would want to deny. The same seems to apply to other good qualities of a man. Intelligence does not require any particular act appropriate to it, nor does wit. But intelligence and

---

19. R.W. Burch: "Are There Moral Experts?" in Monist, 58, 1974, p. 652
20. B. Szabados: "On 'Moral Expertise'" in Canadian Journal of Philosophy, Vol. VII, March 1978, p. 128

courage require that one's behaviour can be described as intelligent or courageous as the case may be. The same seems to apply to teaching. Von Wright would not want to deny that either. There does not seem to be any point in denying that the actions of a brave man should be described as courageous. This is not quite as empty as it may sound. The description of an action depends at least partly on the point of view of the agent, how he saw what he was doing. So the description of an action can never be removed completely from the attitudes of the agent.

Another way of understanding von Wright's words might be that we should be able to have the typical action required by a virtue described in non moral terms. Then we would have the question about what the proper relation was between moral and non moral qualities of an action. The answer is that moral qualities seem to be entailed by non-moral qualities in a way that is described by the concept of supervenience. But I do not think this is what objectors to moral expertise have in mind when quoting von Wright's words.

The idea in von Wright's objection is that any virtue would require a typical action described in the relevant way whenever the occasion arose for showing it, be it courage, justice, charity or whatever. Being good at chess means that you have certain skills you employ when playing that game. There is no sense to be made of saying about somebody that he is good at virtues or morality because there is no comparable skill which can

be employed. This certainly has some intuitive plausibility.

There are two things to notice here. The first, as I have already mentioned, is that this view about the connection between the virtues and action is true of other characteristics as well, like intelligence. To ascribe intelligence to somebody he must regularly exhibit behaviour one can term intelligent, but there is no action typical of intelligence, 'to intelligence'. The second is that the virtues do make a difference to behaviour. There is the most important difference between indiscriminate throwing away of your belongings and benevolence, just as there is a most important difference between courage and foolhardiness or doing whatever you please without any recognition of risk. This difference is important because the benevolent man knows something the other does not. This is what someone who does not know benevolence must learn. It is knowledge of benevolence and other virtues and other good-making characteristics which constitute moral knowledge and make moral expertise possible.

The two objections described above to von Wright's view of the connection between the virtues and actions should now be clearer. They were basically that moral dilemmas made von Wright's view implausible and that moral virtues required some skills. The most important skill of somebody who is virtuous or has moral expertise is that he is able to discern the difference between right and wrong. I have earlier given as an attempt at a

description of moral that it is the ability to do the right thing for the right reason in the appropriate circumstances. So what the two objections come down to is that there is a skill accompanying the virtues and it is the skill of telling the difference between right and wrong and the willingness to do something about it.

It is important to remember that this skill does not exhaust our set of moral abilities. The love of the good must also be included and moral emotions or sentiments. But the skill is a part of it. This skill relies on different non-moral abilities as we judge appropriate. Our judgement about the morality or immorality of surrogacy, for example, is certainly not independent of the information we have of what it is and what effects it is likely to have or known to have. Our judgement about the morality of surrogacy is certainly not independent of the knowledge that it might harm the children born in surrogacy. If we did not know that it might, our judgement would be more likely to be mistaken than if we were in possession of the knowledge.

But why does von Wright not think that there is any moral skill? One important reason, it seems to me, can be detected from his example. With the example from chess playing it is initially plausible to think that there is no comparable skill to the skill in playing chess when we think about morality. But it is important to realize that a game like chess is in important respects different from anything in morality. The game of chess is a closed whole in the sense that it is played with two sets of sixteen

men on a surface divided into sixty four squares. There are a finite number of rules, even though the number of playing possibilities does not seem to be numerable. The skill in chess is the ability to play the men around the table with the aim of checkmating the king of the opponent. Morality does not have any limited area which can be described in a similar way to the chess-table and there is not a limited number of rules in morality and it is controversial, to say the least, whether we can discover the aim of morality comparable to the aim of playing chess. This lends credibility to von Wright's argument.

But it should not be inferred from these disanalogies between chess and morality that no moral expertise is possible. The only thing that can be inferred from these disanalogies is that moral expertise operates under different conditions from expertise in chess. Moral expertise consists in the ability to tell right from wrong and everything that goes with it. Some people seem to be better at it than others. If this ability is accompanied by good moral sensibility and wide knowledge of human nature we get moral experts.

The second argument I want to consider at this point goes something like this. Moral virtues are accessible to everybody. Hence there can be no moral expertise.

The moral dignity of simple people has been an important source of some doctrines in moral theory. Kant, for example, thought the conscience of ordinary people to be a shining example to moral theory and the

sophistication usually accompanying it. But reverence for the conscience of ordinary people can be taken to extremes. You do not have to deny a profound respect for the moral feeling of ordinary people, even though you accept that this feeling can have its limits.

It is true that every rational, human being can be, and probably most often is, moral. There is no need to deny the possibility of amoralism, nor do I need to explain this possibility in this context. It is part of a good upbringing to learn how to deal with certain problems and how to respond to some central examples of evil or goodness. For some people who want to live a quiet, simple life based on some plain, homespun rules it is sufficient to adopt fairly simple rules and a fairly simple moral sensibility too. But for anyone who encounters a serious, moral dilemma these homespun truths will not suffice. For anyone who is actively thinking about radically new, moral problems these truths will not be enough. They may go a long way, and I suspect that they go a longer way with more people than many moral reformers believe. The point is that most, if not all, rational human beings must take serious moral decisions at some stage in their lives and they do not have any other reliable guide than their own thinking and their knowledge of right and wrong. In this respect every rational human being is a moral expert. But it is also true that some people use more of their time in thinking about moral problems than others and the answers of some of them prove more reliable than of others. These people

would be the moral experts. The relation between experts and laymen in this area is not like that between experts and laymen in civil engineering where one can assume that many, if not most, of the laymen know very little about the building of bridges or houses, for example. Most people know a good deal about taking moral decisions and are familiar with many of the views about at least some of the major moral problems. I guess most people have a view about the rights or the wrongs of abortion or infanticide, but much fewer would have a view about the rightness or wrongness of area bombing by the Allies in the Second World War. As Mary Warnock pointed out, family life is central to practically everybody's life and hence views on the major moral, decisions that everybody has to take in the course of his life concerning his family is familiar to almost everybody. So some moral problems will be familiar to most people and some will not. But the relation between moral experts and moral laymen is like that between physicists and people like Stephen Hawking who are at the forefront of his scientific discipline. They have solid knowledge of many of the facts and are in position to dispute many of the inferences and conclusions of the experts.

There is another difference as well. Moral expertise consists in wider knowledge than one can expect in the layman, as I have already indicated. The other difference, no less important, is that moral experts are better equipped to give reasons for their moral views than laymen. It is often the case that people hold moral

views without having a clear idea about how to justify them and often give reasons that are not relevant or not potent in the context. But moral expertise is knowing how to justify your views and recognizing the logical connections between views on different problems and between different arguments. One obvious instance is to recognize the logical connection between views you have on experiments on embryos and your views on abortion. It is, at least prima facie, strange to be willing to accept abortion let us say until the foetus is 20 weeks old but allowing experiments on embryos only until the second week after the fertilization of the egg. So with moral expertise one should expect wider knowledge of what is right and wrong and one should also expect that the moral views can be backed with better arguments than with laymen.

The third and last argument against moral expertise I want to look at in this section is that morality is not teachable and therefore no moral expertise is possible. Usually the important reason given for this argument is that there are no moral academies, no schools where you can go and learn morality.

There are three points relevant here, at least. The first is that moral expertise covers many different areas. It can consist in the ability to clarify moral issues, the ability to live well, the ability to give moral advice and the ability to teach morality to name

some of the abilities that make up moral expertise.[21] It is not necessary for all these abilities to be present to be able to ascribe moral expertise. So it is possible to be a moral expert without being able to live well. It is also possible to be an expert at moral teaching without being able to clarify moral issues very well, although that may seem unlikely. The second point is that there are schools which explicitly aim at teaching their pupils how to live a moral life. The teachers in such schools attempt to impart to their pupils the knowledge that there is a difference between a moral life and a successful one, that even though one was not successful one could live a good and moral life. This is what religious schools try to do. The third point is that there are moral teachers. I guess parents are the most likely ones for children, but there are peers and there are moral teachers like Jesus Christ. This should not be surprising. If moral knowledge is widespread, then we would expect that most people could possibly be moral teachers, at least in some areas even if not in all. We are learning about morality in our everyday dealing with other people. This also explains why there is no need for schools for morality in the sense of schools where we can learn to live well.

---

21. B. Szabados: op. cit. 120 This is not an exhaustive list of the factors in moral expertise.

4.4. Is Practical Ethics Applied Theory?

It should come as no surprise that I think that moral philosophers are moral experts. I have mentioned this before in this chapter. This does not mean that moral philosophers are the only moral experts. There are others like social workers, priests, anyone who regularly has to deal with serious moral problems. Some moral philosophers try to avoid normative problems altogether and deal only with metaethical problems. They would, presumably, not want to be included in the set of moral experts. But those moral philosophers who spend much of their time thinking about normative problems, should be accepted as moral experts. This is not to say that they lead a blameless life or live well. They may or may not do that. But the moral expertise one can expect from them is not of that type. They are educated to argue about moral problems, formulate premisses and conclusions, put forward theories that attempt to explain what is to count as right and what as wrong. Therefore, their special contribution as moral experts is not in leading especially worthy lives or consoling those in distress. I guess many, maybe most, are not equipped to do that. But they are well equipped to articulate moral problems, clarify them, add perspectives to the arguments about them and to advise about them. They are fully acquainted with the major theories about morality and can identify the sort of argument being used in debates. This can

often be important because of the conceptual issues involved.

One natural suggestion about the nature of the moral expertise of moral philosophers is that they are applying moral theories when dealing with moral problems. This is sometimes called the fruits-of-theory model.[22] In putting it like this there are at least three things assumed: 1. that there is an identifiable body of moral theory; 2. that there is a set of problems identifiable as moral problems; 3. that moral theory can be applied to this set of problems. These are important assumptions which can be questioned.

Let us take the second assumption first. There are two questions about it. One, what are the criteria of moral problems. In general moral problems are those which arise in our relations with other sentient beings, the rightness and wrongness of actions affecting them. As should be clear from this, there is no sharp line to be drawn between moral problems and other problems. Two, we could also ask who identifies moral problems. Is it the philosopher? Is it the professional who must face and tackle the problem? Is there anybody in general who can decide authoritatively which problems are moral and which problems are not? Lurking behind this question is the suspicion of many professionals that philosophers think of themselves as final authorities in all questions moral. Are philosophers such authorities?

---

22. J.M. Brown: "On Applying Ethics" in Moral Philosophy and Contemporary Problems, J.D.G. Evans (ed.), Cambridge, Cambridge University Press, 1987, p. 81

To answer the last question first, I think I have already indicated that philosophers are no such authorities and they would be making a serious mistake if they thought of themselves as such. More reasons for this opinion will follow. Is there anyone who has the authoritative say about what counts as a moral problem? I do not think there is. But is there anyone who should have such an authoritative say? That is a different question and a more difficult one. Let us look at the relation between a doctor and a patient. Let us assume that the doctor has to decide between two methods of treatment of cancer, either of which has some major drawbacks and also major advantages. This could obviously be a purely technical decision. But if one of the treatments led to the patient being in a relatively good condition for a limited time and then foreseeably dying but the other leading to the patient living for a much longer time but foreseeably being in much poorer condition and then dying, which one should the doctor choose? This seems to be a moral choice and it inevitably involves consideration of how the patient himself views and values the options, whether he finds it more desirable to live longer where the quality of life is less or to have a shorter life with better quality.

The doctor in question has to weigh the evidence, appreciate the patient's view and the arguments for and against. He might think that it is within his power to decide this or he might take the view that the patient should decide. He might also think that ultimately it is

his conscience as a doctor that is decisive and no further appeal possible. But as before, this is not the case. There is another possibility. It is that his decision is subject to the scrutiny of others, e.g. moral philosophers, who can constructively contribute to an informed and rational decision. Moral philosophers can make a considerable contribution to such a decision by clarifying the issues, adding a perspective, thinking through the problem with the doctor. In any such discussion the philosophical contribution is not the knowledge of facts, the doctor would supply the facts, but in seeing what the possible arguments and inferences are. Sometimes such assistance is not needed, maybe more often than not, but in many difficult and complex decisions the moral philosophers would be able to make a contribution. The decision resulting from such a process would be more informed and, hopefully, more rational than, if the doctor had made it on his own.

There is a strong feeling among many professionals that moral decisions are private in the sense that their own consciences are ultimately the arbiters and it is nobody's business to interfere with that. This has historical roots in Protestantism, I would suggest, and does not survive close reflection. So neither professionals nor moral philosophers are ultimate authorities in moral matters, but their contribution to rational decisions in moral matters can be valuable. Nor is it the case that moral philosophers only take over moral problems as presented by the professional in

question. In his discussion with the doctor one of the legitimate things a philosopher can and should do is to question the description and assumptions made when the problem is approached. There is no single group that can authoritatively identify moral problems and no group should have that authority. It would be unwise of philosophers to accept moral problems as presented by professionals without question because it "risks proceeding with unreflective acceptance of established categories and labels."[23]

Now to the first assumption. The first assumption was that there was an identifiable body of moral theory. This is subject to some obvious counterarguments. I shall mention two. One, there is an identifiable body of moral theories disagreeing among themselves and this throws serious doubt on the assumption that there is an identifiable body of theories. Two, maybe moral theory is simply inappropriate for moral reflection. Maybe moral considerations do not lend themselves readily to theoretical considerations. I shall look at these arguments in turn.

It is true that there are some distinct types of theories of ethics: utilitarianism, Kantianism, Aristotelianism and some others. These theories disagree among themselves about important issues such as the nature of goodness, whether rightness is based on goodness and various other things. But moral disagreement

23. O. O'Neill: "The Power of Example" in Constructions of Reason, Cambridge, Cambridge University Press, 1989, p. 179

need not be theoretical. We can have disagreement about morality among people uninfluenced by moral theory. Opinions differ about many moral issues in most, maybe all, societies. Should children care for their ageing parents until the end? is one example of an issue people disagree about nowadays. Should the rich, Westernized societies distribute some of their wealth to the poorer societies in the Third World? is another. But disagreement can also be at the level of theory. When thinking about morality, we start with views, concepts and practices from our own society. There is no other starting place. But these views can be changed and revised to form a set of considered moral views. This set we can attempt to make a consistent whole. We would expect that at this level disagreement would not be as frequent, vociferous and fundamental as at the level of ordinary opinion. This may or may not be true. In so far as realism, as here argued for, creates the expectation that there should not be fundamental disagreement among moral philosophers, it is disappointing to experience the present state of affairs in moral theory. But we should not infer that this precludes the possibility of moral expertise in any meaningful sense. Many other kinds of experts disagree among themselves and nobody infers that their expertise is any less for it. Also, this means that moral expertise might preside in moral philosophers as a whole rather than individually.[24] There might also be

24. T.C. McConnell: "Objectivity and Moral Expertise" in Canadian Journal of Philosophy, Vol. XIV, Nr. 2, 1984, p. 208

realist explanations for why there is theoretical disagreement among moral philosophers.[25]

The second question or counterargument was that maybe theory was not appropriate for moral thinking. Maybe something in its nature is such that theory somehow distorts moral reality. Hence, the first assumption is wrong and the fruits of theory model misleading. B. Williams argues against moral theory and says that this does not mean that prejudice is the only other alternative. [26] He thinks that the fact that a moral concept belongs to a moral theory gives it no special status and must be justified in terms of our ordinary moral concepts. Two considerations might, though, give credibility to the role of theory. One is that some moral reflection is an attempt to understand our ethical life. This seems, plausibly, to lead to a theory. But Williams thinks that such a theory cannot be a moral theory. The second is that in giving reasons for our actions we are trying to justify them in terms of our practices. But what justifies our practices? This might call on a theory. But the problem is, according to Williams, that this seems to be a claim for a foundation of such practices. Foundationalism of this sort, he argues, is rejected by most philosophers today.

It should be clear that the last point is no problem. I have already made it a part of my account of moral knowledge that foundationalism is implausible. The

---

25. D.O. Brink: op. cit. p.204-209
26. B. Williams: Ethics and the Limits of Philosophy, London, Fontana, 1985, ch.6, esp. p. 111-119

first relies too heavily on the supposed contrast between moral theories and scientific theories. There is no reason to accept it. I argued earlier that Williams' theory of moral knowledge was seriously flawed and, when properly looked at, there was no fundamental difference between moral knowledge and scientific knowledge. The same seems to apply to moral theories and scientific theories. So, I think that there is no reason to believe that somehow moral life does not lend itself to theoretical reflection in the same way as other parts of reality. To this extent, the first assumption is not problematic.

The third assumption was that moral theory can be applied to moral problems. I have argued that there are moral problems, even though no group has a monopoly in identifying them. Also, I have argued that there are moral theories and these theories are trying to explain our moral experience and ethical life. It seems to follow from this that moral theories can be applied to moral problems. But we need to be careful about such applications. It depends on the theory being applied what solution we might get to a moral problem. The solutions of different theories can easily conflict. But this is no argument against the possibility of the application of moral theories. Applying a moral theory is not simply seeing whether a moral property falls under a particular concept.

I want to emphasize that even though moral theories can be applied to moral problems, this does not exhaust

moral reflection or our moral life. For an action to be moral it has to be done for the right reason. A part of being able to act morally is to love the good, to have the right concern and the right attitude. Another part is to be able to think and reason rationally about the aims of our actions and about the means to those aims. To be able to do that we need knowledge and competence in understanding and applying moral concepts and considerations. I think it can be inferred from this that we cannot act morally accidentally or in ignorance of the nature or consequences of our actions. Moral expertise is based on the cognitive part of the ability to act morally. But this type of expertise could perfectly well play a part in forming the right attitude or concern, but this is a complex and long process for any human being and takes place for most of them from infancy until the late teens or the twenties.

4.5. Conclusion

In this chapter I have argued that moral expertise is both possible and plausible. Moral expertise is related to wisdom and can be justified along the same lines, i. e. that reflective actions are more likely to be successful than unreflective ones. Moral expertise contributes to our understanding of morality and towards moral action. There are moral experts and they can enrich our moral lives and assist laymen to come to rational conclusions about moral issues be they social or

personal. Normative philosophers can be moral experts. But they have no monopoly on moral authority or ultimate answers to moral questions and problems. There does not seem to be any reason to reject the possibility of moral theory.

Chapter 5

Paternalism

In this chapter I will examine how paternalism should be understood and justified. Paternalism is making somebody act in his own interest. It does not seem to be ruled out as immoral. I will take a look at the justification of paternalism, how prevention of evil and lack of rationality possibly justify paternalism. Any theory of paternalism must take into account that there are clear cases of paternalism that are justified and also others that are unjustified.

The discussion of paternalism falls into two parts. In thinking about paternalism two assumptions seem plausible. The first is that normal rational beings in full possession of their faculties are able to decide and control their own lives. What is more, we think that they ought to control their own lives. But, secondly, we also think that there are clear instances of justifiable interventions in such decisions for these persons' own good. In this chapter I will discuss the reasons for thinking that there are such instances.

5.1. What Is Paternalism?

To get a grip on the notion of paternalism it is probably best to start with some examples of it and try

to see what they show. There is in this country, and many others, an obligation on motorists to wear seat belts while driving on pain of paying a fine, if caught while driving without one. This obligation has been justified on the grounds that it is in the drivers' best interests to wear seat belts. This is an instance of paternalism.

Another example of paternalism is stopping a man who is about to go on to an unsafe bridge not knowing that it is dangerous. The reason one could give for stopping him is that it is in the man's own interest to know what he is doing. There is a ban in most countries on selling hard drugs to anyone who desires them. One reason given for such a ban could be that it can be in no one's interest to use such drugs on himself. Hence the ban is meant to protect his own good. When a parent stops his own child eating sweets on the grounds that it is bad for the child his action is paternalistic just like the ban on narcotics.

All these examples seem to be examples of paternalism. They have two things in common. First, all paternalistic actions, laws etc. seem to be interferences with the actions of somebody different from the interferer. Second, these interferences aim at the good of those interfered with.

Both these things need some explaining. One question about the first part is whether it is possible to be paternalistic toward yourself, i.e. can you interfere with your own action for your own good. In general we can describe paternalistic actions like this: x makes y do A

for his own good. The question raised by paternalism towards yourself, or self-paternalism, is whether it is possible that x and y are one and the same person. This means that it is not a necessary feature of a paternalistic action that it involves two agents. I think it could be reasonably believed that self-paternalism is an instance of paternalism proper. This may seem incoherent but I think it is not and that it is worthwhile to try to spell out the sense in it.[1]

It is claimed that the central example of this phenomenon is Odysseus and the Sirens. Odysseus wanted to hear the song of the Sirens but he was not confident in his ability to withstand the seductive lure of their song so he asked his crew to tie him to the ship's mast and refuse any request or order to set him free while their song lasted. The crew put wax in their ears to prevent them from hearing the song. The claim about this example is that Odysseus interfered with his future actions by his request to the crew, his own will prevented him from doing what he desired to do while the Sirens' song lasted. This interference with his later desires is claimed to be self-paternalism.

It might be pointed out that on this description it is the crew's action that is paternalistic, not Odysseus' request. But this would be to miss the point. The idea is

---

1. Husak, D. N.: "Paternalism and Autonomy", in Philosophy and Public Affairs, Vol. 10, nr. 1, 1980, p. 43-46 discusses such examples and argues that they amount to paternalism towards oneself, which is his term, or self-paternalism, a term suggested to me by John Skorupski.

that the request is paternalistic because it limits Odysseus' liberty during the Sirens' song. Other examples of self-paternalism can be when one fails to apply for a credit card because of the fear that one will abuse it when obtained. Or when one makes an irrevocable promise of giving away one's future inheritance. Both these would count as self-paternalism.

There is one legitimate question which should be asked about these cases. What if you change your mind about your promise, contract or commitment when it is about to be fulfilled? This is not a problem for ordinary contracts and promises because there is another party who can, and in most cases will, hold you to your contract and promise. If it could be argued that such holding to one's promises was in one's own interest, this might count as paternalism proper. But what about commitments like not applying for a credit card or giving away your inheritance? If you change your mind about them, your earlier decision is revoked. You may have reasons for your change of mind like thinking that you are now a better judge of your financial position than you were earlier or that giving away your inheritance is silly. You may be right about this or wrong. But the point is that this is a perfectly normal thing to do and there is no stopping you. Now, if the original decision is so easily revocable, then there does not seem to be anything to hold you to your earlier decision. And what is more there does not seem to be any sense to be made of

usurping one's own choice as must be the case, if there is any sense in self-paternalism.

But I think we should look at the example of Odysseus again. The crew refuse his order to release him at his request. It is true to say that Odysseus request is self-paternalistic and the crew's refusal is paternalistic. But why do we say it is paternalistic? It is paternalistic because Odysseus' new choice is denied and the crew discount his new belief about what is for his own good. Odysseus original request is self-paternalistic for essentially the same reason. It discounts his new decision and request because Odysseus accepts the belief that his judgement will be so affected by the Sirens' song that his desires and choices during that time will be entirely unreasonable or irrational. His fatal desire to swim to the Sirens which the song produces is irrational. What Odysseus is doing is to exclude the possibility that he will act on his own desires while the effects of the song last. His request to the crew is just instrumental in preventing him in acting on his irrational desires.

It is perfectly possible to attempt to prevent yourself from acting on your future desires without having anyone else ensuring that your will is enacted. If you have a key to a Pandora's Box and you want to make sure that in the future you will never be tempted to open it, then you would have the key destroyed. It is also possible to imagine cases where one might have a drug which had the effect of, let us say, your desire for

alcohol never arising. Should we inject ourselves with this drug? If we do, this is a case of self-paternalism. If one asks for sterilization, then one is making sure that in future one can never beget children. One may have perfectly good reasons for this, like poverty. These cases have in common that the decisions are not revocable. So even though it is true that some cases of self-paternalism are such that the decisions to be self-paternalistic are rather easily revocable, this is not true in all such cases. So this fact cannot be counted against the possibility of self-paternalism. Self-paternalistic decisions are meant to be binding on oneself and some are revocable and some are not. So self-paternalistic decisions really do usurp one's future choice.

There is another feature of paternalistic actions and self-paternalistic ones. It is that it seems to be a necessary feature of any paternalistic act that it is believed to be superior to the choice which is denied. In most cases, it seems, that the superiority is based on better knowledge and better judgement. One might avoid drinking alcohol because one believed that the influence of it on one's judgement would make the judgement worse. Odysseus knows that under the influence of the Sirens' song he will become irrational. One believes that getting hold of a credit card will make the temptation to buy many things one cannot afford irresistible. The irrationality of the judgement and the irresistibility of the temptation are marks of a flawed judgement. This

indicates that presumed superior judgement and presumed superior knowledge is an assumption of any paternalistic action. This should be carefully distinguished from disrespect and indignity. It seems to be possible to be paternalistic without showing any disrespect to or demean the one who is the object of such an action. It is not demeaning to oneself to avoid drinking alcohol, but it is based on the knowledge that my judgement under its influence will be inferior to my judgement when sober. In the same way the crew of Odysseus' ship is not in any way showing him disrespect when not obeying his orders while the Sirens sing. They are doing what they know is in his interest and what he really wants.[2]

It is sometimes argued that paternalism is nowadays a pejorative term and always used as a criticism of whatever it is applied to. Now if this is taken to mean that paternalism is always or necessarily unjustified, this is false. This probably comes with the idea that paternalism is somehow necessarily demeaning for the one who suffers it. I do not doubt that this is a widespread idea. I sometimes get the feeling that it is treated as a self-evident truth that every human being is the best judge about anything concerning his own interest and

---

2. D. N. Husak uses his notion of "paternalism towards oneself" to argue that the relation of superiority-inferiority is not a necessary feature of paternalism and he takes this to indicate that paternalism is not opposed to autonomy. I think he is right about the relation between autonomy and paternalism and self-paternalism, but, as should be clear from what I have said, I think he is wrong about superiority and paternalism. It seems to me that he confuses superiority with disrespect and indignity.

hence that, somehow, paternalism is an insult to his dignity and always unjustified. Hence, it might be claimed it is always applied as a term of criticism and a pejorative term. But this is false. Some of the examples I cited at the beginning seem to me to be uncontroversially justified, especially when a parent prevents his child from eating sweets. If we say that it is paternalistic when a child is refused a sweet, it seems to me that we are not criticizing it. We are only expressing the fact that the parent is denying his child's choice and preventing it from doing something which ultimately is not in its own interest. The same applies to the action of the crew of Odysseus' ship. There simply is no sense in which it is disrespectful or demeaning not to set Odysseus free or denying a child a sweet. Hence it seems to be false to say that it is a part of the meaning of the term 'paternalism' that it is pejorative.

There can be one-party and two-party cases of paternalism.[3] This should not be confused with paternalism towards oneself discussed above. The defining characteristic of one party case is that the action

3. Feinberg, J.: The Moral Limits of The Criminal Law Volume 3: Harm to Self, New York, Oxford University Press, 1986, p. 9-10. See also G. Dworkin: "Paternalism", in Laslett, P. and Fishkin, J. (eds.): Philosophy, Politics and Society, Oxford, Basil Blackwell, 1979, p. 81-82 where he distinguishes between 'pure' and 'impure' cases of paternalism. Feinberg wants to use 'direct' and 'indirect' paternalism. This distinction does not exactly coincide with the distinction between one-party and two-party cases of paternalism, even though in practice most, maybe all, two-party cases will be treated as indirect paternalism and one-party cases as direct paternalism. This refinement does not matter for my argument.

prevented, interfered with or prohibited only concerns yourself. A ban on suicides, for example, or self-mutilation or the use of drugs is a one-party case. The action banned is only performed by you on yourself. A two-party case occurs when the action interfered with, prevented or banned is requested, wanted or asked by another party. Even though the ban is for the good of the requester it does not ban the request but the action of the one who carries out the request. This is what happens when the sale of drugs is banned or when it is banned to lend money at, let us say, the rate of 50% p.a. Such a ban is for the good of the drug user or the borrower but limits the liberty of the drug seller and the money lender. Two-party cases add complications to the discussion of paternalism, but they should not influence the case for or against paternalism. I will have one-party cases in mind when discussing it, unless I say otherwise.

Earlier when I described the first examples of paternalism I said they seemed to have two things in common. One was that they seemed to show paternalism to need one agent to interfere or prevent another agent's action. I argued that this is not a necessary feature of paternalism. The second thing was that this interference was done for the good of the agent himself, it was in his own interest. We need to examine the notion of good, or interest briefly at this point.

The notion of interest or good in this context can be divided into two. On the one hand interest includes

benefits. The provision of health care is an instance of benefit provided by the state in Western states nowadays: it is meant to enable the members of a society to enjoy the benefit of health and lead a life not marred by too much pain. On the other hand when referring to the interests of a person we can be talking about harms and the question would be how we should prevent a harm from befalling him. This prevention would be intended to protect his good. This seems to be the more frequent case occurring in discussions of paternalism. But harms can be of two kinds. First, harms can be a setback to a person's interest wrongfully inflicted. This basically is harm not consented to. If I decide to take part in an experiment with drugs, for example, am informed about the possible consequences of taking it and the danger I am putting myself in, then I consent to taking the drug by indicating in some conventionally accepted way that I will take part. If something goes wrong and I am injured, then I am not harmed in the sense of a setback to interests unconsented to. But I am harmed in a second sense. Harm in this second sense is injury. In this sense you are harmed independently of whether you consented to the action that harms you. Let us say that you take up smoking without realizing how harmful it can be. Before you become addicted somebody points out to you the dire consequences that can accrue to you if you smoke regularly for a long time. Until this was pointed out to you, smoking will have harmed you in both the senses described. But if you decide to go on with smoking after

you have been informed, then you can only be harmed in the second sense. Both these senses of harm can be at work in arguments about paternalism.

It is important to remember that paternalism does not apply to all interferences with somebody's action. It only applies to those interferences aimed at the good of the person interfered with. The denial of his choice, which is an inevitable part of paternalism, is based on what his interferer takes to be better knowledge of the interferee's good. The good can either be some benefit or a prevention of harm. Obviously these assumptions can be mistaken and what was believed to be for the good actually turns out to be bad or even a disaster. If the state decides, as it did in the U. S. in the twenties, to prohibit the sale of alcohol to its inhabitants, this can be done on the ground that it is in the best interest of the inhabitants not to consume alcohol at all, that they ought to be teetotallers. The banning of the sale would be believed to be the best way of achieving total abstinence from alcohol by the population. This is a two party-case of paternalism. But as it turned out in the U. S., the prohibition was a disaster and apparently did not have the effects desired. This should not really have surprised anyone, because such a ban cannot have been based on any thorough understanding of human nature nor on deep understanding of the type of society in which it was enforced. Similar experiences in other countries did produce similar results, even though they were not as spectacular failures as in America. The explanation of

this disaster seems to me to be not the fact that the prohibition law was paternalistic but that it was based on poor knowledge and understanding. So there is a difference between an action or law being paternalistic and this action achieving its professed aim. An action can be paternalistic, even though it does not actually promote my good, if my good was intended.

It is important to remember that the operative notion of harm in this context is injury. The reason for this is that the good of agents is objective as I have argued for before. This means that it is perfectly possible to be harmed without realizing it. The consequence of this is that wrongful harm in the sense of harm unconsented to plays a less important role. What is important is to find out what is the nature of the values involved and what actions can possibly be harmful.

The good of persons in question as I conceive of it is welfare in a very general sense. This really amounts to happiness, if happiness is not understood as a blissful state of mind but as goodness of life consequent upon our decisions and conduct in the course of our lives. This good is not limited to pleasure and the absence of pain but it includes health and wealth, for instance, and self-determination because self-determination is a value of which we can have more or less. One of the values aimed at in a paternalistic intervention can be self-determination incoherent though this may sound.

There is a distinction drawn between hard and soft paternalism.[4] To throw light on this distinction it is well to remember one of the examples above of a paternalistic action. The example was of preventing someone from going on to an unsafe bridge. What was left unsaid in that example was what we would do, if the person going out on to the bridge insisted on going after having been informed of the danger. This illustrates the difference between soft and hard paternalism. Soft paternalism holds that it is justifiable to prevent someone from performing an action, possibly harmful to himself, so long as it is reasonably clear that the agent's action is not fully voluntary. If he does not know that the bridge is dangerous, it is fully justified to stop him and inform him. His action cannot be fully voluntary in so far as he is not fully informed about what he is doing. It is reasonable to suppose that he does not want to put his life in danger by going out on to an unsafe bridge but in all probability he wants to cross the bridge. If this is true, his action is not voluntary.

But let us suppose that the man listens to what we have to say, understands it and yet he insists on going out on to the bridge. Soft paternalism holds that it is only justified to prevent his action in so far as he is not fully informed and his action not fully voluntary. So a soft paternalist would believe himself in the right in stopping the man from going on to the bridge. But he

4. Feinberg, J.: op. cit. pp. 12-16

would refuse to hold him back after imparting his information and making sure that the agent had understood what was being said. He would believe that it was nobody's business to interfere with a voluntary action. A hard paternalist, on the other hand, would not accept this as a constraint on his interference. He would hold that he was fully justified in preventing someone from doing something that was harmful to himself. The hard paternalist would not believe that this applied to any action harmful to the agent himself, but only those actions that would result in more injury and damage than the prevention itself. In assessing the consequences of the intervention the hard paternalist would have to take into account the pain and discomfort caused to the agent interfered with. This is really just a part of saying that the intervention should be in the interest of the agent interfered with.

There is a connection between soft paternalism and the notion of harm as described above which deserves to be pointed out. The soft paternalist attempts to prevent harm in the sense of setback to interests unconsented to, i. e. wrongful harm. But he does not concern himself with self-inflicted damage or injury which is fully voluntary. That is out of bounds, so to speak. Soft paternalism shirks interfering with voluntary actions of rational beings. This avoidance shows it for what it is: it is anti-paternalism. This should not be a surprise to anyone because it is admitted by those advocating soft-paternalism and take it to be one of the virtues of soft

paternalism.[5] It is obvious that soft paternalism covers some very important considerations concerning paternalism but when it comes to the crunch, it shies away from preventing damage to an adult for his own good against his wishes. But this is where we enter the area of paternalism proper. Successful soft paternalism draws a boundary around the personal sphere arguing that no invasion into that area is justified. Hard paternalism or paternalism proper holds that such invasions or interferences for the person's own good are sometimes justified.

It seems to me that the most powerful reason for accepting soft paternalism, or more accurately anti-paternalism, rather than paternalism proper is the belief that all paternalistic interferences are coercive. Many infer from this that paternalism is morally repellent though this is unwarranted unless you add more assumptions. But are all paternalistic interventions coercive? There is reason to believe that they are not. Let us consider an example.

> "Mr. N, a member of a religious sect that does not believe in blood transfusion, is involved in a serious automobile accident and loses a large amount of blood. On arriving at the hospital, he is still conscious and informs the doctor of his views on blood transfusion. Immediately thereafter he faints from loss of blood. The doctor believes that if Mr.

5. Feinberg, J.: op. cit. p. p. 15

N is not given a transfusion he will die. Thereupon, while Mr. N is still unconscious, the doctor arranges for and carries out the blood transfusion."[6]

It is quite clear from this example that the action of the doctor is paternalistic and yet at the same time it is not coercive in any obvious sense. There is no attempt to control the behaviour of the patient or to interfere with his liberty. Another example showing the same characteristics is of a mother on her deathbed asking for her son. The doctor knows that the son has just been killed attempting to escape from prison. But he tells the mother that he is doing well. The doctor certainly lied to the mother but he did not coerce her in any way or control her behaviour but his action was paternalistic towards her.

This does not show that no paternalistic action involves coercion. But it shows that some do not. So coercion is not a necessary feature of paternalism and at least in that respect it need not be morally repellent. When thinking about paternalism, especially legal paternalism, it is plausible to assume that coercion is a necessary feature of paternalism. Enforcing paternalistic laws, like banning riding a motorcycle without a protective helmet, seems necessarily to involve coercion.

---

6. Gert, B. and Culver M. "Paternalistic Behavior" in Cohen, M., Nagel, T. and Scanlon T.: Medicine and Moral Philosophy, Princeton, Princeton University Press, 1981, p.202

Many cases of paternalism that soft paternalism endorses, like stopping somebody crossing an unsafe bridge, seem to involve coercion. But it is possible to have paternalism without this feature so it cannot be a necessary part of it. So when objecting to coercion you need not necessarily be objecting to paternalism.

For a paternalistic act to occur it must be possible to identify one's good independently of the good of others. If this is not possible, paternalism becomes impossible. In a society like any Western society, and Western is meant to include Japan and Australia as well as Western Europe and U.S.A., in the last two or three decades, it has been possible to identify somebody's interest or good independently of the interest of others. In these societies paternalism has been possible. But it is easy to imagine a society where everyone's interests become so intertwined that it is hardly possible to separate them. In a society at war, for example, where the contribution of everybody is valuable and non-contribution might be harmful to the effort of the whole society, then the identification of individual interests as independent of the interests of others becomes difficult. The problem is that anything affecting one's own interests affects those of others as well. This also seems to apply to small, cohesive societies living in difficult physical surroundings like Inuit societies. In such societies no action seems to be entirely for my own good because all my actions affect or may affect the good of others. So it seems that no purely paternalistic

action is possible in such a society. This does not mean that somebody else cannot decide something for me irrespective of my own wishes but he cannot do it for my own good because my own good is so closely connected with the good of others. These considerations only show how paternalism is relative to normal conditions in society.

I have argued for at least three necessary conditions that any action must fulfill to be paternalistic. 1. When A is paternalistic towards B, he decides what to do for B's good which B would otherwise not have done. In the case of the doctor above he decides what to do for B's good contradicting B's expressed wish. This is a case of noncoercive paternalism, as I argued, but it is paternalism and it would count as manipulation. 2. A intends his decision to affect, directly or indirectly, B's good. It is a fact that some, maybe many, paternalistic actions are not successful. This does not change the fact that they are paternalistic. 3. A is willing to discount B's belief that A is not promoting his good. This is what is meant by accepting superiority as a feature of paternalism.[7]

These three conditions are necessary for any action to be described as paternalistic. I do not want to claim

---

7. Archard, D. argues for a tripartite definition of paternalism in his "Paternalism Defined", Analysis, Vol. 50, No. 1, January 1990, p. 36-42 similar to what I have offered. But it is a necessary part of his definition that B's choice is denied and his opportunities diminished. As will become clear, this cannot be a necessary feature of a paternalistic interference. Banning duelling, for instance, actually increases the opportunities of those who are subject to the ban. I offer my conditions as necessary but not as a full flown definition in view of the discussion above.

that they are sufficient and, hence, a full blown definition. But it deserves to be pointed out that any action described as paternalistic has to be caused by those beliefs described in the three conditions mentioned above to be properly paternalistic.

5.2. The Justification of Paternalism

In the first part of this chapter I have argued for a particular set of necessary conditions for paternalism. This is meant to cover all cases of paternalism. These cases can either be justified or unjustified depending on the case being discussed. There do seem to be unproblematic cases of both kinds. In this part of the chapter I want to discuss how paternalism can be justified and what the distinguishing line between justified and unjustified instances of paternalism is.

There is a a very old ban on duelling in this and other European countries at least. Such a ban seems to be an instance of a paternalistic law because it can be justified by the good of those who participate in duelling. This law may have had and may have now other justifications than the interests of the duellers themselves and hence have a mixed justification. But that need not concern us here. Let us assume four things. First, that everyone in a particular society where duelling is being banned prefers "most of all not to be

confronted with dueling situations".[8] Second, everyone prefers to preserve his honour in the conventional way if and when a duelling occasion arises. Third, a legal ban on duelling would prevent any such occasion to rise. Fourth, all the persons concerned have no other relevant desires concerning the merits of duelling regulations. From these assumptions, it is claimed, it follows that "a legal ban against dueling would be nonpaternalistic, since nobody's freedom is being restricted against his will."[9] If this is correct, then conditions like these change a law from a paternalistic one into an unpaternalistic one.

What happens here? First, it is clear that everyone wants the ban and, second, the ban successfully achieves its aim of preventing any duelling occasion from arising. Given these assumptions the question of paternalism does not arise, except for one thing. It is the second assumption which might possibly raise the question of paternalism. It is that everyone would defend his honour in the traditional way if the occasion arose. But this desire does not have the same priority as the desire to be rid of the practice of duelling. So even though these desires seem to be contradictory they are not because the second has priority over the first one if they come into conflict. If we look at the definition of paternalism, a law banning duelling given these assumptions, cannot, it seems, be paternalistic. Such law does not supplant

8. Arneson, R. J.: "Mill versus Paternalism", in Ethics, Vol. 90, Nr. 4, 1980, p. 471
9. Arneson, R. J.: op. cit. p. 471

anybody's choice, it does not discount anybody's belief about his own good, but it does affect the good of the members of this imagined society. But that has no consequence whatsoever for the paternalistic nature of such laws, be they of such a nature. I think the reason given for most, if not all, laws is that they promote the general interest. In this respect the ban on duelling is no different from any other law.

It was claimed above that the ban on duelling was nonpaternalistic because "nobody's freedom is being restricted against his will." Or to be more accurate against his most important desires because it can be argued that the ban could be against one of his less important desires. But the important thing to notice about this example is that the assumptions are extremely strong. As soon as we move to more plausible assumptions, such as that most people might prefer a ban on duelling, the problem of paternalism returns. What about those who would not? And what if the opinions of some of those who prefer the ban were based on irrational grounds? Should we accept such preferences? Does the irrationality affect the paternalism-nonpaternalism issue? These questions are a hint of some of the problems that would have to be faced in a thorough discussion of these issues. These assumptions about the desires of the agents involved can be used as justifications of paternalism and in that way the justification can influence whether an act is paternalistic or not.

There is one justification open to someone who wanted to accept paternalism about duelling given the assumption that most people preferred to avoid the danger of the practice. It is true that the ban on duelling limits liberty in the society in which the ban holds but such a ban can also increase opportunity for these same members. Duelling could be a threat encouraging men to use their time and energy to keep up their skills and strength in fighting with a sword or pistol or whatever weapon the tradition endorses. Even though one accepted a high regard for valour and physical strength, one could understand and desire other valuable things. These other things might include love of family, wealth and health. The social practice of duelling as a way to settle quarrels or differences of opinion or conflicting claims to the same goods, could be a drain on one's time and energy. In a society where duelling is practised you must develop your skills or endanger your life and property, unless you are prepared to give in to many or most demands. So when this practice is banned you have many more opportunities to pursue the other things you find valuable. So this limit on your liberty can easily increase your opportunities. This would justify paternalism, assuming that most people want to avoid duelling.

This justification cannot be generalized to any practice in a society which someone might find dangerous or undesirable or even to practices many or most might find dangerous or undesirable. But it can certainly be

applied to some other things. The same considerations can apply to wearing protective motorcycle helmets but they cannot be applied to religious worship, it seems. Wearing a protective helmet can save your life and saving life certainly creates more opportunities than dying which closes all opportunities. In banning a particular religion there is no obvious sense in which this ban can open up opportunities for those suffering the ban.

The basic assumption in discussing paternalism and its justification is that a conflict can arise between one's values. As I hope is clear from the discussion above about paternalistic acts their object is to promote the good of the person concerned. Paternalism aims at the person's happiness or well-being. If well-being were the only value or the highest value of all in a person's good, this would be unproblematic. But there are other values which have the same claim on our attention and effort. One of them is self-determination. In some matters at least we find it desirable that the person himself decides what he should do rather than someone else. It is fairly clear that some such decisions have worse consequences than decisions by another person better able to judge the situation. When such situations arise the values of self-determination and well-being conflict. When there is a conflict there is often also a need to decide which value is more important. It is not always possible to find a satisfactory way to settle such conflicts.

It is not only the fact that these two values can conflict which makes a justification of paternalistic actions possible. If we have reason to believe that the agent himself can assess his whole situation and come to a conclusion about whether he ought to seek the advice or help of somebody else then there is no reason to interfere with his action and the question of paternalism does not arise. So in addition to a conflict of values there must be reason to believe about an agent that his decision in some particular action of his is encumbered[10] for the possibility of justifying paternalism to come about. In saying that a decision is encumbered we are saying that there is reason to believe that the agent does not consider his interest as well as he would have done if his decision had been unencumbered.

It is important to distinguish between two things in this context to avoid confusion. One is privacy, the personal sphere or actions concerning oneself, the other is one's best judgement of one's own interest or one's own good. It is perfectly possible to accept that the personal sphere ought to be up to each individual himself but at the same time deny that everyone is the best judge of his own interest under all circumstances. It is a plausible consequence of this denial that it ought not to be the case that everyone is the ultimate authority on his own interest. So when there is reason to believe that a decision is encumbered and the agent's judgement not as

10. This is the word J. D. Hodson uses in his article "The Principle of Paternalism" in American Philosophical Quarterly, Vol. 14, Nr. 1, January 1977, p. 65

good as it might be, this does not necessarily imply anything about the personal sphere. It might be better all things considered to leave the decision to the agent. But it is a possible justification of a paternalistic action.

When we say that a decision is encumbered this means that we are comparing two courses of action and one of them is better than the other and the agent has chosen the worse one. In such a situation the question can be about the benefit to the agent. Problems about the benefit to the agent are sometimes grouped under the label of perfectionism[11] rather than paternalism and the latter limited to preventing the harm that might befall the agent if he pursued the course of action he decides on himself. This distinction is not important for my discussion. The justification of paternalistic actions becomes possible when there is reason to believe that the decision was encumbered and the action might be harmful to the agent.

The encumbrances can be of different kinds. They could be irrationality, recklessness, negligence, incapacity, youth, senility, emotional stress, difficult, coercive circumstances and so on. If we had reason to believe that any of these encumbrances were operative on the judgement of the agent we could be justified in preventing him from doing what is harmful to him. But to

11. Joel Feinberg talks about legal perfectionism when discussing this beneficial function of the law in The Moral Limits of The Criminal Law Volume 4: Harmless Wrongdoing, New York, Oxford University Press, 1990, ch. 33, pp. 277-317

be justified it is necessary to compare the harm that follows the proposed or decided course of action and the cost that might accompany the interference. One important part of assessing the possible harm of the interference is the resentment the agent is likely to feel because of the interference. But there are others like the effect on the agent's self-esteem, the probability of the interference achieving the goal it set out to achieve. These separate things influence the justification of the paternalistic act. The nature of the encumbrances will, for example, influence the evaluation of the paternalistic act achieving its aim. If we are dealing with a mentally retarded person, it is less likely that a paternalistic interference will cause severe resentment and more likely that it will achieve its aim, especially because the responses of the mentally retarded are often easier to predict than the responses of fully mature adults.

When contemplating a paternalistic interference there seem to be three possible factors on which the justification depends: 1) the comparative value of the intervention and the action prevented; 2) the encumbrance decreasing the rationality; 3) the consent of the agent. If the aim of the paternalistic intervention is of relatively low value, let us say securing good manners, then the value of the intervention is in all probability outweighed by the value of having the agent himself deciding on the course of action. There are different kinds of encumbrances and they affect the agent and his

judgement in various ways. There are obvious differences, for example, between somebody who is mentally ill and another agent who is reckless. The reckless agent does not care about the consequences of his action. In so far as those consequences hurt others we do not think of an intervention as paternalistic, but in so far as they hurt the agent himself an intervention preventing such harm would be paternalistic. Someone who is mentally ill very often cannot decide rationally what is in his interest and what to do in a particular situation. Often he cannot appreciate what he is actually doing nor the consequences of his action. This does not necessarily mean that it is simpler to intervene in his behaviour.

The third factor is the possibility of consent. In many, perhaps most, cases of paternalistic intervention there is no possibility of consent prior to the intervention. So the possibility and probability of subsequent consent becomes crucial to the justification of a paternalistic act. Odysseus gave prior consent to his being tied to the mast. So his crew were justified in keeping him stuck there. If they had untied him, then that would have been an unjustified paternalistic act. It would have been unjustified and paternalistic because it would have discounted Odysseus' voluntary and rational decision, gone against his explicit consent and harmed his interest. But prior consent to a paternalistic action is the exception from the rule which is that prior or

contemporary consent is virtually impossible. So subsequent consent often seems the only possibility.[12]

There are two things that need to be discussed about consent in the context of paternalism. The first is what sense to make of consent in general and subsequent consent in particular. Two, I will outline an argument why consent has no place in a philosophical theory of paternalism like the one I have argued for.

Consent can be either actual or possible and it can be prior or subsequent. This means that there are four categories to consider. The first is actual prior consent. The second is possible prior consent. The third is actual subsequent consent. The fourth is possible subsequent consent.

It is certainly open to us to use any of these four notions of consent as a constraint on the justifiability of paternalism. The strongest notion is actual prior consent. This is what happened in the example of Odysseus. Prior to the interference of the crew he consented to their action. I take it that this is extremely rare and if it were a necessary condition of a justifiable paternalistic intervention, then they would be practically ruled out. But the problem with actual

---

12. This account of the role of consent is contradicted in "Justifying Paternalism" by Gert, B. and Culver, C. M. in Ethics, vol. 89, 1979, p. 200 they say: "The definition also makes clear that one feature used in justifying a violation of a moral rule is not open to A when acting paternalistically, namely, the consent of S toward whom A is violating the rule. For if A has S's consent, A's behavior is no longer paternalistic." The example of Odysseus shows how it is possible to have consent and yet be paternalistic.

prior consent is that it does not really make sense to require it. If the prior consent had not been expressed because of some irrelevant difficulties, then this would be crucial about the justification of a paternalistic action. Now we might introduce a requirement on all these categories that they required relevantly full information. So if an actual prior consent was not forthcoming for some irrelevant reason we would have to assess whether it would have been forthcoming in the absence of this irrelevant reason. This has the effect of weakening the constraints in all these categories on paternalistic actions.

The notion of subsequent consent is not very clear. Is there, for example, a time limit to it? Can a meaningful consent be given to any act after the event? Can you consent to a paternalistic intervention long after the event? The basic problem with the notion of subsequent consent is what difference it makes to the original action. It is difficult to see that it makes any difference at all. If a man rapes a woman, for example, and a week later the woman comes to the conclusion that she ought to have consented to it, this does not in any way change the nature of the original act. And in this context it seems awkward, to say the least, to speak about subsequent consent; even paradoxical. The reason is that the notion of consent does not seem to be related in the same way to all acts. Some acts are evil because they are not consented to. The example of rape is typical in this respect. Sexual intercourse consented to is just

sexual intercourse. Sexual intercourse unconsented to is rape. That is why subsequent consent is so inappropriate for such an action. It simply has no place in that context. But if it is a question of an act that secured benefits or prevented harms consent has no effect on the nature of the action and the action does not change in nature depending on the consent.

The approach of the soft paternalist to the question of consent is that prior voluntary consent is the only acceptable possibility. But lacking that the limit to paternalistic actions would be provided by the possibility of the removal of the encumbrances on voluntary consent. If somebody intended to kill himself, for example, we would be justified in stopping him to ensure that he understood what he was doing and if he did not, to make him wait until he calmed down and regained his senses and his balanced judgement. But if he still wanted to do away with himself after reflecting calmly on his proposed course of action, then the soft paternalist would not feel justified in stopping him. The reason is that if the agent decides after calm reflection to go on with his action, then paternalistic intervention is not admissible according to the soft-paternalist. He rejects the notion of subsequent consent and argues that all consent must be voluntary actual consent. So subsequent consent carries no weight for the soft paternalist and is a counterfeit notion of consent as Feinberg argues.[13]

---

13. Feinberg, J.: The Moral Limits of The Criminal Law Volume 3: Harm to Self, New York, Oxford University

This does not deter some who are willing to consider themselves soft-paternalists from using the notion of subsequent consent.[14] This notion is then supposed to apply to all cases of prima facie justifiable paternalism. If the notion of subsequent consent has a role to play, the possibilities for paternalism increase.

I said above that there were two things that needed discussion in connection with consent. I have finished the first which is to see what sense one can make of the notion of consent in this context. Now I turn to the second part to find out if the notion has any part to play in a philosophical theory of paternalism.

The first thing to notice about these considerations concerning subsequent consent is that the notion itself seems peculiarly inappropriate. What is in question is whether an agent can reasonably come to appreciate after an intervention that it was the best thing to do, even if his own choice was supplanted by the interferer's. In this context it is more appropriate to speak about endorsement or ratification than subsequent consent. These concepts do not have the contradictory ring about them that subsequent consent has and they do seem to catch everything that is important about subsequent consent. But this is only a terminological point.

Press, 1986, pp. 182-183. Feinberg uses the example of rape to argue that subsequent consent is contradictory.

14. See Carter, R. : "Justifying Paternalism" in Canadian Journal of Philosophy, Vol. VII, Nr. 1, March 1977, pp. 133-145 where the author argues that subsequent consent is a perfectly legitimate part of justifying paternalistic actions.

I have already explained that I understand a person's good as objective. This means that it is not constituted by his expressed preferences and it is possible to be harmed without realizing it. I understand happiness and welfare very generally. A part of this happiness is obviously to enjoy reasonable health and not be deprived of the necessities of life. But more importantly for my purposes self-determination is a part of welfare. Self-determination is a value and we can have more or less of it and in general the more the better. It may have to be balanced against other things obviously, but that is no problem. This is what we would expect about every value at some stage or another. So if it can be argued that a paternalistic intervention increases the agent's welfare in relevantly important ways and the disvalue of overriding the will of the agent is not too great, then it is justified to do so. If we have reason to believe that the agent's will was encumbered, then this decreases the value of the will. The value of the intervention outweighs the value of the will more easily in that case. This is all we need to have a justification of a paternalistic action.

Consent of whatever type cannot be a deciding factor on this account. If we are contemplating a paternalistic intervention, then consent does not seem to have a large role to play in our theoretical considerations. The relevant considerations are only the values that ought to be increased against the possible disvalue and the rationality of the decision. If the action is justified,

then the value aimed at outweighs the value of the course of action prevented. This is the conclusion of D. H. Regan when he discusses the example of Odysseus and his crew:

> "If what I have just said is correct, the command is neither necessary nor sufficient to justify the crew's decision to keep Odysseus aboard. What justifies their decision in the standard version of the tale is the fact that swimming to the Sirens is understood to be an irrational choice, with fatal consequences, made under preternatural compulsion."[15]

It is important to emphasize to prevent misunderstanding that I do not want to deny that consent can and should have an important role to play in the actual carrying out of paternalism. For an official to have the consent of the subject can be crucial in justifying the action. The only thing I want to say is that in theoretical deliberations on paternalism we ought to think of the values concerned. One of the most important ones is autonomy. This ought to bring in all the relevant considerations about the agent and his own role.

The conclusion to this is that there are only two relevant considerations in determining justification of a

---

15. Regan's article "Freedom, Identity and Commitment" in Sartorius, R. (ed.): Paternalism, Minneapolis, University of Minnesota Press, 1983, p. 128

paternalistic action. The first is the value of the interference, the second is the rationality of the choice or decision supplanted.

5.3. Examples of Paternalism

There are other cases of apparently justified paternalism which soft paternalism has difficulty in dealing with. It seems to be true that most reasonable people would be willing to accept limits on access to narcotics. Limits on access to narcotics can be justified by reference to harm to others. But there seem to be clear cases where the only justification to be had for a ban on free access to narcotics is paternalistic, i.e. it can only be justified in terms of the good of the agent.

I want to describe three more cases where the soft-paternalist seems to have difficulty or at least a case can be made for a hard-paternalistic justification of a ban on certain practices or of forcing agents to adopt certain practices. I do not know whether there is a ban on brainwashing, but if there were it seems to be possible to make a case in paternalistic terms in a similar fashion to the ban on free access to narcotics. Any consent to undergo brainwashing is problematic because the decision is irrevocable. After undergoing brainwashing you cannot reflectively assess the original decision. The same applies to the addict. He cannot reflect on his condition rationally because the addiction distorts his judgement to such an extent that he is not

his own master. Even though some narcotics do not cause addiction after a single injection, it is safe to assume that one injection of heroin, at least, causes addiction.[16] In that sense the decision to try it is irrevocable. The ban on unlimited access to narcotics prevents you from implementing a decision to try. The irrevocability of such an implementation seems to be a good reason for the ban. The soft paternalist has to hold that even if the decision is irrevocable, the agent should be allowed to implement it if it is quite clear that his decision is voluntary, i.e. is not based on any mistakes about facts or his own values. The soft paternalist would believe that a test for voluntariness ought to be strict in this instance because of the irrevocability, but presuming that the agent had passed it, it is not justified to stop him. But why ought we to feel justified in allowing the agent to harm himself in such a drastic fashion? And why ought the agent to feel justified in doing so? The hard paternalist is not obviously wrong to want to stop the agent from harming himself nor is the soft paternalist obviously right.

There are two more examples I want to discuss. One example is of the obligation to wear seatbelts while driving a car. I think this obligation is enforced by law in many, maybe most, Western countries. The argument

16. Strangely enough, morphine, which one would assume, caused addiction just as strongly and readily as heroin, is not assumed to do so when administered to hospital patients suffering from severe post-operative pain. The danger of addiction is said to be a "totally unjustified fear". See D. Doyle: "Pain Control" in The Practitioner, 22 March 1990, Vol. 234, p. 285

usually given for this law is that it makes it more likely that drivers and passengers in a car will survive if they have an accident. Obviously much depends on whether the factual assumption is true, but I think it is reasonable to believe it to be true. The soft-paternalist would be willing to educate drivers in the dangers they face on the roads and to make sure that they understand these dangers. But going further than that is not justifiable according to him if the grounds for doing so are the agent's own good. So he would have no way of stopping a driver who was not willing to wear his seatbelt and realized all the dangers and the increased likelihood of surviving them if the seatbelt was worn. The hard or real paternalist would have no serious qualms forcing him to wear his seatbelt as the legislators have done.

The other example is of forcing motorcyclists to wear protective helmets by law. The reason for forcing them to do that is that the helmets increase the probability of surviving serious accidents. There is good reason to believe that safety helmets make all the difference when a serious accident occurs.[17] Death in such accidents results in a large majority of cases from head injuries. Protective helmets prevent head injuries. So helmets make it more likely that motorcyclists survive serious accidents. It is obviously in the interests of the motorcyclists themselves to wear them. That is the

17. Feinberg, J.: The Moral Limits of The Criminal Law Volume 3: Harm to Self, New York, Oxford University Press, 1986, p. 135

hard-paternalist's reason for forcing them to wear the things. But, as before, the soft-paternalist would not be willing to go further than ensuring that all motorcyclists are knowledgeable of the facts both cognitively and emotionally. But should it be left ultimately up to the motorcyclist to wear a safety helmet or not? Or is it justifiable to force him to do so?

In these last two examples there is a real question about whether the paternalistic intervention is a wrong suffered by the agents. It is at least questionable whether the paternalist is justified in forcing the driver to wear a seatbelt and the motorcyclist to wear a protective helmet. Any agent wanting to drive a car or a motorcycle might reasonably ask why it is not and should not be up to him to decide whether he wore a helmet or a seatbelt. The risk in these activities is not much greater than in some other activities we are unwilling to limit in any way, like mountain climbing. So he might appeal to the law to protect him from such interferences which the agent who wanted to inject himself with heroin cannot apparently do. The argument the hard-paternalist could use at this point is that these two types of limits to liberty are trivial in any normal life. It simply does not matter all that much to any reasonable man whether he uses a seatbelt or a protective helmet. In comparison to the good such practices can secure in the event of an accident the triviality becomes even more prominent. So considerations of the value of the consequences seem to determine that this is justifiable and also

considerations based on rationality. A rational agent would choose the better alternative of the two or more available to him. An additional reason for accepting a ban on not wearing protective helmets is that in the majority of cases where the driver survives the accident when not wearing a helmet he regrets his omission. When he wore his helmet he tends to be glad that he did.

This is an argument one needs to be careful about. It is not implied by it that the more trivial the action forced upon the unwilling agent the more justifiable it is other things being equal. There is a limit to such justifiability. Whether I wear a brown jacket or black is something which I ought to decide. It simply does not make sense to try to justify forcing me to wear a brown rather than a black jacket. My good is not involved in the decision to wear a brown jacket rather than a black one. So forcing on me either of these possibilities cannot be justified in any way, paternalistic or any other. If it were not a question of forcing or coercing but of deciding for me something which might affect, directly or indirectly, my good, then it might be justified because coercion is possibly more harmful than merely a decision for me. The manner of the paternalistic intervention is clearly important. But up to a point it seems to hold that the more trivial a pursuit is, the more justified it is paternalistically to intervene, assuming that the intervention secures an important good. Beyond the point where the intervention possibly affects my good, there is no justification for intervening.

But the opposite does not hold. It is not true that the more important or central to one's life an action is which is being interfered with, the less justification there is for intervening even assuming that the good secured is critical. In some cases one might even think that because the decision the agent is about to take is so central to his life, it is possibly justified to interfere. But I believe such cases to be rare and the intervention always questionable. If a friend was about to marry a woman you knew to be unsuitable for him and that an unhappy marriage was likely to depress him severely, then you might think it justified to interfere. But precisely because a decision as important as this one is so central to the agent's life, then this is also a reason against interfering. Let us assume that your friend can easily become depressed and if he gets seriously depressed, he might become suicidal. This had happened in the past and all the indications are that it might happen again. Let us also suppose that you know, that the wife to be of your friend had lied to him about her past. You know, for example, that she was for a time a prostitute. You know also that your friend is sensitive to the past of his wife to be. What are you to do? If you decide to tell your friend that his wife to be was a prostitute at one time, then in all probability you have prevented him from getting married and you have behaved paternalistically towards him. He did not ask for any such information and he might even be unwilling to listen. If you do not do anything before he marries, your

friend might discover after the marriage that his wife had a disreputable past. This could trigger his depression and make him suicidal.

This would obviously be a difficult decision and there is no clear answer to what is the right thing to do. To that extent a paternalistic intervention in circumstances of this type is always questionable and I do not think it possible to formulate any general rule about such interventions in such circumstances. But it is clear that prima facie at least a paternalistic intervention is not ruled out and might just as well be the right thing to do as leaving things alone. Precisely because there is so much at stake you might be justified in intervening.

So the weight of the values of the proposed action and the value of the course of events following the intervention have to be assessed and judged. There is no simple relation between the two. Adding the value or disvalue of the intervention itself makes it even more complicated. No general conclusions should be drawn from the fact that wearing seatbelts and protective helmets can be trivial compared to other things in one's life.

The first thing to notice about all of these examples except one is the absence of encumbrances on the decision being prevented. In the case of education there is youth or immaturity which is the encumbrance and the avoidance of education is the evil being prevented by forcing the child to go through education.

It is quite clear that it is possible that a conflict between self-determination and well-being can occur. When someone decides to try a narcotic like heroin it is obvious that this decision will harm him. In such cases there is a conflict between self-determination and happiness or well-being and it is precisely in these cases that the possibility of a justified paternalistic intervention is present. If one takes the view that self-determination always overrides well-being there is no problem of paternalism. But there is no obvious reason to ascribe such importance to self-determination.

One reason for thinking paternalism problematic is that paternalism is widely believed to be a pejorative term. I have discussed this earlier in 5.1. and rejected it. But there might be one reason for this belief that I have not discussed. We might accept that a paternalistic action always broke some moral rule.[18] Hence, we might conclude that paternalism was pejorative. The main question in this context is whether a paternalistic act always violates a moral rule.

It is quite clear that many paternalistic acts violate a moral rule. It is also clear that many justifiable paternalistic acts can violate a moral rule. So the distinction between justifiable and unjustifiable paternalistic acts does not seem to coincide with the

---

18. See Gert, B. and Culver, C. M. : "Paternalistic Behavior" in Cohen, M., Nagel, T. and Scanlon, T. : Medicine and Moral Philosophy, Princeton, Princeton University Press, 1981, p. 205; and Gert, B. and Culver, C. M. : "The Justification of Paternalism" in Ethics, Vol. 89, 1979-80, p. 199

distinction between those acts that violate a moral rule and those that do not. When drivers are obliged to wear seatbelts or motorcyclists are forced to put on protective helmets, then it is arguable, at least, that they are being deprived of a part of their freedom. Depriving someone of his freedom is violating a moral rule. Banning unlimited access to narcotics or prohibiting brainwashing are not clear cases of any violations of moral rules, unless one assumes that prior to any paternalistic act there is a right to any action one voluntarily wants to perform.

It should be noted that the right to non-interference cannot justify permission in these areas. Supposing that one's having a right to non-interference depends on a belief that in some areas at least self-determination is more appropriate than accepting the decision of someone else, perhaps because each individual is the best judge in these areas. But this is exactly the problem. Whatever view we take of the nature of the right to non-interference we might believe that paternalism could be justified. The reason for saying that these are not clear cases of violations of moral rules is that there is no obvious argument for saying that self-determination has priority in these cases because of the irrevocable injury involved. The only sense to be made of the claim that these cases involve violations of moral rules is to assume that there is a rule saying that agents have rights of self-determination to any action whatsoever.

In the example discussed earlier of a doctor injecting blood into a patient who had made it clear prior to the operation that he did not want any new blood into his body I concluded that this did not constitute an instance of coercion. Hence coercion is not a necessary feature of paternalism. But the action of the doctor seems to be a violation of a moral rule. The patient had expressly forbidden the doctor to inject new blood into his body. The doctor's action is paternalistic because it is against the wishes of the patient and aimed at his good. It might possibly be said that, had the doctor not known the patient's will, his action would not have been a violation of a moral rule. But when the patient had expressly forbidden it, I cannot see it as a non-violation. This obviously depends on how the moral rule in question is expressed. It might go something like this: rational beings have unqualified sovereignty over their bodies when in full possession of their faculties. This sovereignty in those areas he has expressed his will about extend over those periods in time that he is unconscious and not in full possession of his faculties. The doctor's action, then, must be a violation of a moral rule.

Curiously Gert and Culver argue that the doctor's action in this case is not a violation of any moral rule but they think that his action requires the doctor to break a moral rule later on when the patient wakes up.[19]

---

19. Gert, B. and Culver, C. M. : "Paternalistic Behavior", op. cit. p. 207

Then he either must lie to the patient or cause severe pain to him because of his action. Therefore they think that any paternalistic action involves violation of moral rules. But I think that if there is any violation of a moral rule in this case it is because the doctor's action is against the expressed wish of the patient. This is clearly a manipulation and a violation of the patient's autonomy. They seem to believe that there is no violation involved in the action of the doctor because it is not coercive. But this is false. A violation of a moral rule can occur without any coercion.

It is obviously important in this context to have at least some idea about what a moral rule is and hence what amounts to a violation of it. In this respect Gert and Culver are not of great help. They say in one place: "It is not necessary that one explicitly hold some theory about what counts as a violation of a moral rule. All that is required is that one believes A is doing one of the following: killing; causing pain (physical or mental); disabling; depriving of freedom, opportunity or pleasure; deceiving; breaking a promise; or cheating. All of these are universally regarded as requiring moral justification and hence are regarded by us as violations of moral rules."[20] The first question about this is how do we know that these are instances of violations of moral rules? Their answer is that because these are regarded as requiring moral justification they are

---

20. Gert, B. and Culver, C. M. : "The Justification of Paternalism", op. cit. p. 200

violations. The assumption must be that all and only those actions regarded as requiring moral justification are violations of moral rules. But this must be false. The class of actions requiring moral justification is much larger than the class of actions that are violations of moral rules. Supererogatory actions, where one does more than is morally required, seem to need moral justification. But they are not violations of any moral rule. The problem is with what we are saying when talking about an action requiring moral justification. When justifying an action morally we are giving our moral reason for it. But this we can do with any action and it does not seem to be limited to those actions that can be treated as violations of moral rules. So it seems to me that it is not a feature of paternalistic actions or their justifications that they are necessarily violations of moral rules. But they certainly need moral justification.

One suggestion about justification is that it must be universal.[21] This is Gert and Culver's idea. In their discussion universal means two things. On the one hand it means that if all rational beings could agree to a certain action it is right and hence justifiable. On the other hand it means that if a particular agent in particular circumstances is to be justified in doing X, then everyone in relatively identical circumstances must be justified in doing X. When they say that for any

---

21. Gert, B. and Culver, C. M. : "The Justification of Paternalism", p. 200-206

justification to be valid it must be possible to make it public, this only applies, it seems, to the first of these two meanings, i.e. to make it public is to put it to other rational beings. This is what they say: "...he must be willing to claim that all rational persons could agree with his judgement that in these circumstances Mr. K. should be deprived of his freedom for a limited period of time."[22] As should be clear, universality in this sense is perfectly possible without universality in the second meaning obtaining. This means that all rational beings could agree to this particular deprivation of freedom without implying anything about all other beings in relatively identical circumstances.

Gert and Culver say also:

> "What is then needed is to determine whether to publicly advocate this kind of violation. This is done by deciding whether the evil prevented (relieved) or avoided by universally allowing this kind of violation outweighs the evil that would be caused by universally allowing it. If all rational persons would agree that the evil prevented by universally allowing the violation would be greater than the evil caused by universally allowing it, the violation is strongly justified; if none would, it is unjustified. If there is disagreement, we call it

22. Gert, B. and Culver, C. M. : "The Justification of Paternalism", op. cit. p. 201

a weakly justified violation, and whether it should be allowed is a matter for decision."

In this quote both the senses of universality are at work. In this context 'allowing the violation' means 'allowing paternalistic intervention'. The relations between the two senses of universality can be put schematically like this:

1.	all agree	all are allowed
2.	all agree	one is allowed
3.	all agree	none is allowed
4.	some agree	all are allowed
5.	some agree	one is allowed
6.	some agree	none is allowed

In this schema when I say 'all agree' this is short for all rational beings or persons. When I say 'all are allowed' this means that all persons in relevantly identical circumstances are allowed to intervene paternalistically. When it says 'none is allowed' this means that the paternalistic intervention is prohibited, i.e. the person about to be interfered with should be free to do what he wants. I have left out in the left column the possibility when no one agrees because this is the extreme possibility of everybody having his own opinion and not agreeing with anyone else.

---

23. Gert, B. and Culver, C. M. : "The Justification of Paternalism", op. cit. p. 203

The first thing which is obvious about this schema is that none of these six possibilities described is incoherent. None of them seems to contain a contradiction. So we might opt for any of them, if we found some plausible argument for accepting it. Gert and Culver think that the first possibility, nr. 1, represents strong justification for a paternalistic action, that is when all agree that all are allowed. When they say that a paternalistic action is unjustified, 'if none would' the unstated word is apparently 'agree'. So it is unjustified when none agrees that everybody is allowed. This is one of the possibilities that I dropped because it seems to mean that everybody has an opinion and it is different from everybody else's. This does not seem to me to make sense. What they are probably after is that there might be some who disagreed very strongly but strength is something which is left out of the schema. Or they might be saying that even though some agreed most would disagree. Let us accept this interpretation for 'none agrees'.

But notice that what they are saying is that this represents an unjustified act:

7. none agrees all are allowed

This is a much weaker unjustifiability than is to be found in the 3rd possibility where all agree that none is allowed, i.e. all agree that it is prohibited. This seems to me to be the strongest sense in which a paternalistic

action is prohibited or unjustified. The 6th possibility only means that there is disagreement about the prohibition which implies that there are some who agree. But the fact is that it is prohibited and there is disagreement about it.

The main question to be asked about this schema is how strong should the universality requirement be on a justified paternalistic action. It seems to me that possibility 2 is all that is required. This needs some comment. If we understood the first part literally it would mean that no paternalistic action would be justified unless all agreed. This is clearly far too strong. If just one disagreed because of misunderstanding or bigotry of some sort or another, then this would disqualify a paternalistic action from being strongly justified. So the interpretation of this would have to be all who are fully rational and either reasonably informed or fully informed. This shows the difference between the two parts of these interpretative possibilities. The first part is really an epistemic statement, the second is an ontological one. It seems to me that a paternalistic action might correctly be allowed, even though only some, or even just one, recognized it as such. But I think it is all right to have a stricter requirement on the agreement because of the dangers inherent in paternalistic actions. That is why I propose the second possibility as a description of a strongly justified paternalistic action.

But why should we accept that only one was necessary for a paternalistic action to be justified? I do not want to exclude the possibility that sometimes, maybe most of the time, we would have to weigh the consequences of allowing all to interfere paternalistically. But it seems to me that there might be cases where it could be justified to allow a person to interfere paternalistically without weighing the universal consequences. One reason for accepting the possibility that there is only one who is allowed in some situation to act paternalistically is that there is no algorithm for applying moral rules. The application itself cannot be guided by rules. So it seems to be possible that there is only one who is allowed. Another reason for saying this is that one factor we have to take into account is the actual values of the person being interfered with. A third reason for accepting this is that we would not be violating the universality requirement. What we would have in such a case would be a specific principle about the justifiability of a paternalistic interference. A specific principle can be universal. Generality is contrary to specificity, not to universality in the ontological sense.[24] These three reasons seem to support the claim that only one is necessary for a paternalistic action to be justified.

One objection to this might be that it is just a part of meaning that an action is allowed that anyone in

---

24. Hare, R. M.: Moral Thinking, Oxford, Clarendon Press, 1981, p. 41

relevantly identical circumstances ought to be allowed as well. But I do not think it is warranted to take this as a part of the meaning of the term. But it certainly seems to me correct to say that if there is a moral reason for allowing this, then this reason is a reason for any moral agent in relevantly similar circumstances. This seems to me to depend on the nature of rationality and moral objectivity. It is also true that even if applying rules is not universalizable even though the rules themselves are universalizable, this does not affect the nature of the rules. They can be as general as we like. In giving reasons for a decision for a paternalistic intervention the reason would obviously have to appeal to some general ideas to be convincing. But this does not exclude the possibility of the action being only appropriate in this instance.

All things considered it seems to me that for a paternalistic action to be strongly justified all who are rational and reasonably informed must agree that at least one is allowed to act in that way.

5.4. Rational Desire

At this point I want to emphasize that I do not propose to reject Gert and Curver's theory of the justification of paternalism altogether. It seems to me to have some parts that are correct, but there are these two I have discussed which seem to me either wrong, that paternalistic acts always break moral rules, or unclear,

the condition of universality. They hold that there are three things relevant in specifying a justification of a paternalistic action. The first is the moral rules violated. I have rejected this. The second is the evil prevented or benefit caused and the possible evil caused by the interference. I have already indicated that this is part of assessing the justifiability of a paternalistic act. The condition of universality is a part of such an assessment. The third is the rational desire of the person affected by the paternalistic interference. These last two are parts of any justification of paternalism, it seems to me.

This third part needs some elaboration. What is rationality and what is a rational desire? I have already argued for a way of understanding reason, value and rational desire in 2.3. Let us look at two examples. The first is the example of blood transfusion to the unconscious patient I have mentioned before. The doctor injected blood into the patient even though he had expressly forbidden it. It seems clear that when the patient wakes up and the doctor tells him the truth that it will cause pain to the patient. But this pain is consequent upon the patient's belief about the blood of others and upon his expressed prohibition of injecting blood into his body. He believes that another's blood is an impure substance and that getting it into his body will damage him beyond redemption or whatever. The salient fact about this belief is that it is false. There does not seem to be any natural sense in which it can be

explained how transfused blood can be impure. I do not know whether a reasonable theological explanation can be given of these beliefs, but I seriously doubt it. So the request not to be injected with blood is, in this respect at least, irrational. So the question arises whether rationality can justify the doctor in his paternalistic action.

Let us look at another example. A patient is in a critical condition after a serious accident. He was driving in a car with his family and everybody in the car but him died in the accident. Now he regains consciousness and asks how the others fared. It is clear to the doctor that the patient's life depends on the reply he gets. If he gets the unvarnished truth, he could easily deteriorate and die. If the doctor lies and says they are fine, the patient might recover and ultimately resume living his life again. What should the doctor do? I believe most people would say that the doctor ought to lie to the patient accepting that this is a paternalistic act and he might ultimately have to face up to what he has done, if the patient recovers.

I take it that the paternalism in the first example is problematic. It is problematic at least in the sense that there is disagreement about what the doctor ought to do. In the second example I would not expect there to be serious disagreement. This is especially in view of the fact that I do not suppose that the deception of the patient would go on indefinitely. It seems to me that the difference between the responses to these two examples

would be explained by the fact that we assume that the patient in the second example wants to live and that this is a rational desire. So the doctor in lying to his patient is assisting him to achieve his own goals through a paternalistic action. But in the first example the doctor would be going against the expressed desires of the patient and would not be assisting him to accomplish his own ends even though it was for his own good. The reason most frequently given for this is, I believe, that the doctor would be imposing his values on the patient in the first example but not in the second. This is considered unjustified interference.

In the second example the instrumental view of reason is at work. According to this view every person has some given basic desires and ends and the role of reason is to discover, assess and come to conclusions about the best ways to reach these ends. Reason is not believed to be capable of evaluating the ends themselves, only the means. The affective side of rational beings supplies the ends. This seems to me to be one of the motivations behind the view that the reflective consent of an agent should not be overridden by a paternalistic intervention. In some sense the ends are thought to be up to each and every individual.

But as I have argued earlier in 2.3. this conception of reason is inadequate. It is inadequate because basically it makes good sense to discuss and criticize the aims people accept. Some aims seem to be trivial in a perfectly ordinary sense and an action or even a life

having a different end from the trivial one is more valuable. The value in question seems to be independent of the desires fulfilled and seems also to be independent of whether somebody actually aims at these ends. The values also seem to depend on beliefs just like the belief about the value of impure blood. Sometimes these beliefs can be false. One conclusion one might want to draw is that in those cases where values are in some sense irrational, there is reason to interfere paternalistically and thereby impose values on the irrational agent.

We should be careful about this. I want to make one thing clear. It is that it is not the imposition of values on somebody which is objectionable. When somebody is punished for a crime certain values are being imposed on him. But they are being imposed in the name of the interest of others so to speak. One of the aims of punishment is to attempt to deter the offender from committing a crime again and to deter others as well. This is justified on the grounds that the interests of others demand that people are deterred from violence. For this to work the actual offender is being forced to accept the restraint of the interests of others. Deterrence would not work if he did not accept ultimately that he had to take the interest of others as prior to some of his own tendencies. This is a limited sense of value but it seems to be value all the same. So the imposition of values is not objectionable as such, only if it is done for my own good.

Why are the lives and interests of others so much more important than my own life and interest that I am justified in harming my own life and interests but not those of others? One way of arguing for this is to find a more general reason or rule. How would that reason look? It would go something like this: People are not permitted to harm people. This rule is perfectly general in the sense that it covers both myself and others. If it is correct, then it seems to follow from it that I am not permitted to harm myself any more than I am allowed to harm others. There seem to me two reasons for accepting this rule. One, the fact that a life or interest is the life of another does not seem to be essential to the ban on harming. There is nothing apparently inappropriate in formulating the rule in this more general way. Second, if it is true that the correct way of thinking about rationality is as agent-neutral, then this general way of formulating the rule seems to follow from it. So I think there is no apparent mistake in putting the rule like this. From this I think we can infer that my own interest are just as important as the interests of others. It seems that I should be just as ready to accept paternalistic interventions as interventions based on the interests of others. Why it does not work out quite like this I will try to argue in the next chapter. But I hope these reflections have at least made it plausible that such a general rule is not inappropriate and not implausible.

The notion of rationality covers ends as well as means. When an agent has a mistaken belief about something influencing his values and decisions then it is possible to say that the value and decision is irrational. This gives the notion of irrationality a very wide range and it is certainly not justifiable to infer that a paternalistic intervention is all right if the decision of the agent is based on a false belief. In this wide sense 'irrational' gets to some extent a similar meaning to 'unreasonable'. But to be unreasonable is not the same as being irrational even in this extended sense. Being unreasonable is taking a month's wages at the beginning of a month and spending them all wining and dining. (This may sound more difficult than it actually is.) Or going out on to an unsafe bridge after being told of the danger. These two courses of actions or traits of character would be unreasonable. It seems to me that none of them needs depend on a mistaken judgement although they would probably all be signs of a character fault: self-indulgence, perversion, volatility. One way of looking at this is to say that labelling these actions as unreasonable is just saying that on our scale of worthwhileness they are pretty low down. But this would presumably not give anyone a reason for intervening with such an action.

Yet it is rational for an agent who has these character traits to try to change them. And it is perfectly possible that such an agent might be better off protected against his own choices. But when it is stated

that it is rational for him to change his character it is implied that the character he has is faulty in some respect. I think it would be generally agreed that perversity, volatility and self-indulgence are faults of character. They are faults because they can easily harm you. This fact ought to give any rational agent a reason to attempt to correct the faults. One might want to infer from this that if this is true, then a paternalistic intervention could be justifiable. But the problem with faults of character like these is that the agent himself very often does not see these traits as faults. They are ingrained parts of himself and they do not give him reason to think that he ought to change his ways. This is often the biggest part of the problem.

When deciding whether a paternalistic interference is justified or not we look at the evil prevented by the interference and compare it with the cost that accompanies the intervention. In a case of character faults the cost of preventing an action could easily outweigh the evil of leaving the agent to his own devices. One of the considerations would have to be that it is often very difficult to get an agent to look at his traits as a fault. Just because of that it might very well be that in most cases the agents should be well left alone, even though one accepted the analysis above. But we should also note that in assessing the justification of a paternalistic action it is easier to discount the beliefs of the agent because they are false and unreasonably biased due to the character traits.

There is one more consideration which needs to be looked at. There are some goods which are such that it seems reasonable to suppose that any agent would want them. These are goods or valuable things like health, education, minimum wealth and even life itself. Given the arrangement of modern societies, it seems reasonable to believe that any rational agent would want these things. It applies at least to education, wealth and health that some minimum amount seems to be necessary in modern societies for a decent life. Life itself does not allow any gradualness. Either you have it or you don't. But any rational agent under normal circumstances would want life rather than not. This is obviously not to deny that there might be circumstances where it is reasonable for a rational agent not to want life. When I say that it is reasonable to suppose that any rational agent would want these goods, I do not think that this is only a plausible supposition. It seems to me that this is what a rational agent would and ought to want, if he were fully informed.

The question about these goods is whether they justify paternalistic actions any more than other interests of the agent? It applies to health that it seems to be necessary for anything else we want to do and in this respect it is like life. So any rational agent ought to want his own health and his own life. In the absence of knowledge about the wishes of the agent, one would be justified in securing the agent's health and life. But what about the cases when the interferer knows that the agent wants to do away with his life or damage

his health? Is he justified in intervening? If we look at the case of the doctor who gave his patient blood who had expressly forbidden it he seems to be doing his duty. His duty is to preserve life subject to the limits of when it is probable that the patient's life will be excruciating misery rather than reasonably free of pain. So when somebody believes falsely that blood injected into him is impure, then it seems that the doctor would be failing in his duty if he did not preserve the patient's life.

This is so because the doctor has a specific duty in virtue of his training and position. What about others who are not bound by their professional duty to preserve life or health? Are they justified in going against the specific consent of an agent who prefers to die rather than have blood injected to him? Is the law, for example, justified in obliging a wife or husband of such an agent to take any reasonable action to preserve his or her life and in fault of that to prosecute him or her? The answer to this seems to me to hang on the importance we give to self-determination. I will deal with this in the next chapter.

5.5. Conclusion

I have characterized paternalism as an action aimed at the good of the agent but going against his desires at the time of the interference. The paternalistic interference can either be aimed at some benefit for the agent interfered with or preventing harm befalling him.

The notion of self-paternalism seems to make good sense. The justification of paternalism is based on the comparative values of the course of action prevented, of the intervention and the course of events promoted. The justification is also subject to a condition of universality and the rationality of the agent. The basic assumption in any possible case of paternalism is that there is a conflict between the value of self-determination and well-being.

Chapter 6

Autonomy

In this chapter I intend to examine the notion of autonomy. I will discuss Feinberg's theory of autonomy in some detail and express doubts about it. This requires discussion of his four meanings of autonomy and how they lead to different conceptions of autonomy. After that I look at the nature and value of autonomy and, lastly, I try to say something about the relation between autonomy and paternalism.

6.1. Autonomy and Self-Determination

Before discussing the notion of autonomy I think I better counter an objection to the argument so far. It is that the consequentialist framework argued for in 2.4. and the objective, agent-neutral conception of the good simply rule out any place for autonomy. So autonomy is not a problem for a moral theory making these assumptions. The only relevant question is what has the best consequences. The obligation is to maximize the good. This implies that whenever autonomy and happiness come into conflict we ought to opt for happiness because choosing autonomy would lead to diminished good. The problem with paternalism is precisely that being allowed to decide his action will bring harm or even disaster on

the agent himself. So paternalistic interference with the agent's action seems to follow from the assumptions.

This is too swift. As I have already pointed out both autonomy and happiness are parts of welfare and are valued in common-sense morality. I take it that they are also values in our considered moral intuitions. But having them both seems to be unstable and even incoherent. But there are at least three replies to this worry. The first goes something like this. The agent always has the duty to perform the action that has the best overall consequences. A strong aversion to harming seems clearly to be a part of this duty. It also seems to be a part of this duty to be inclined to prevent oneself and others from harming themselves. But the inclination to prevent others from harming themselves is subject to the constraint of the harm that might come with the interference. A part of this harm might be the resentment of the agent interfered with. The best motivation of a moral agent is an important part of his theory: motivation ought to aim at maximizing the good. This motivation, if not tempered with something like autonomy, could create the temptation to interfere with the action of others for one's own gain. So the best motivation of rational agents might be an inclination to respect autonomy, even though in some situations such an inclination did not produce the best possible consequences. But in the long run such a constraint would produce better overall consequences than no constraint if we view it in the light of realistic human psychology.

A second response to this problem could be this. There are presumptions and aversions that govern our actions. Even though these presumptions conflict with a consequentialist calculation in a particular case and going against a presumption seems to be better than acting according to it, the best thing to do might be to go on accepting the presumption. The argument is that consequentialist reasoning determines what presumptions and aversions or rules to accept and in turn these rules take priority over consequentialist calculations in particular cases. So if there is reason to accept a presumption in favour of autonomy, then this should be followed even when following it might have worse consequences than intervening with the agent's action. This does not mean that autonomy would take priority over welfare come what may. If the consequences of not intervening were really bad, it would be justified to break the rule.

The third reply says that it is better overall to leave every agent to lead his own life as best he can. Most agents develop fairly quickly a good sense of what is in their own interest and what is not. So in general we can accept that every agent is the best judge of his own interest. Accompanying this must be a restraint on others not to interfere with those actions of the agent that primarily concern him. So autonomy ought to be respected, unless the consequences are very severe.

These three responses should be sufficient to suggest that the consequentialist framework I have

accepted does not preclude the problem of autonomy from arising. To this extent at least the argument in this essay is not incoherent.

Autonomy is used in various contexts. The first thing to be said about it is what I will not talk about. We can talk about metaphysical autonomy. This notion comes into play when we discuss the compatibility or incompatibility between freedom and determinism, whether it is possible to say that we decide our actions if at any given moment in time a particular causal situation can only have one conclusion or one outcome. Autonomy in this context refers to our ability to decide actions independently of causal connections. This is not the autonomy I shall be speaking about. We can also talk about epistemological autonomy. Autonomy in this context is to what extent we have chosen ourselves the principles on which our knowledge rests. This is not the autonomy I will be discussing. Autonomy can have other senses beside these and I will discuss some of them below, but these two senses will not be relevant.

Autonomy is the major idea used against the justifiability of paternalism. Accepting autonomy as a legitimate notion of morality with its role to play implies that there are certain limits on paternalism, even that it is unjustifiable altogether. That is how some theoreticians believe autonomy blocks paternalism generally. In what follows I will argue that autonomy places no blanket ban on paternalistic interferences and

seems to be compatible with it in some instances at least.

It seems that the simple basic notion in autonomy is that something is up to me. To say that something is up to me means that I decide it, I control it. There are some obvious instances of this in any normal life: whether I should wear a brown or a black jacket. Another is whether I should fall asleep on my left or my right side. It just makes no sense to say that anyone but I should decide this. The idea that somebody else might do so does not seem incoherent, though. If I were kept under a close enough supervision the supervisor could take these decisions for me. And if he had the means, he could enforce these decisions. But this idea does not really make sense, not because it is incoherent, but because it seems to be very difficult to discover a point to it. What could possibly be the point of making me fall asleep on my left side rather than my right side or wear a brown jacket rather than a black one? This obviously applies only under normal circumstances. As soon as we start talking about abnormal circumstances things change. If my serious backache would only heal when I slept on the left side, then it might make sense to have someone forcing me to sleep on my left side if I proved unwilling to do so myself. If my life depended on wearing a brown jacket rather than a black one at a particular time, then it might make sense to have someone compelling me to wear it. But under normal circumstances this does not apply. So it is the point of the action that seems to determine

whether there could be a sense in making an agent do something against his will.

There is reason to point out that this notion of autonomy as I have described it up to now is very limited. Autonomy seems to say that something is up to me only if it does not affect my interest or good. Even though we add that this applies in normal life and in normal life there are at least some things which do not affect my interest, then this does not help very much and the notion seems to be of limited value. But given that this notion of autonomy makes sense, this seems to me to be the essential idea which is worked on with much subtlety in the writings on autonomy. The slim features of this concept may not look very promising and as it is it will certainly not satisfy many philosophers. But there is no reason to prejudge the concept as described but to look and see if and how it can be extended to incorporate the concerns of those philosophers interested in having it do important work.

The basic question about autonomy as described here, it seems, is where the boundaries of autonomy lie. If autonomy is limited to those areas that do not concern one's good or interest, then it is very limited indeed and it could not serve as a constraint on paternalistic interventions. The reason is that paternalism is by definition an interference for the agent's own good. If autonomy is limited to those areas where no interests come into play then it cannot possibly limit paternalism. It was clear from the example above that as soon as one's

good came into play interference was not pointless. So in those areas important to normal, rational agents such an interference seems to make sense.

I said above that the basic notion of autonomy seemed to be that something was up to me. It seems that threats to this control can come from two sides. They can come from other agents who try to take control of my actions. This can happen through slavery, for example, or when somebody attempts to get me addicted to drugs. On the other hand my control of my decisions and actions can be undermined by my desires which can overwhelm me and make me do things which I do not want to do. It can also be that I do not find the desires that I have desirable but am, for some reason, unable to change them. This idea of control over myself leads us to see that autonomy in this sense is not simple and requires complex abilities to be operative in the agent. Corresponding to the two threats to autonomy it seems possible to discern two main ingredients in it. The first is the ability to withstand overwhelming emotions. This ability can be described as the ability to see and act on good reasons. The second is being free from domination. To be a slave is always being subject to the will of another. The first can be called internal autonomy, the second external autonomy.

Before going further I think it is worthwhile to look at a theory of autonomy which makes autonomy more complex than I have done. Feinberg distinguishes between

four meanings of autonomy.[1] The first meaning is the capacity to govern oneself. This capacity is a natural ability and as such it allows of degrees. Some agents are better at governing themselves than others. There is another notion closely connected to this one. It is the legal notion of competence which is an all or nothing notion. Either you are legally competent or you are not. Both these notions are relevant to our understanding of autonomy. I will discuss them both in the next part of this chapter.

The second concept is the actual condition of self-government. This means that someone who has the ability to decide for himself what to do actually does so. Someone who has the capacity to govern himself may not in fact govern himself. He may be a slave or he may be an unwilling subject of a tyrant. It is only when he actually governs himself that he is autonomous. Autonomy in this sense divides into twelve parts in Feinberg's analysis arousing the suspicion that he is not discussing a single concept but many.[2] There can be many ways of preventing me from attaining autonomy in this sense. A serious illness forcing me into coma, poverty or being overpowered by brute force all prevent me from attaining autonomy. Both the notions of autonomy seem to be at work in these examples. Among the examples Feinberg takes of virtues based on this concept of autonomy are self-

---

1. Feinberg, J.: The Moral Limits of Criminal Law, Volume Three: Harm to Self, New York, Oxford University Press, 1986, pp. 27-51
2. See op. cit.: pp. 31-44

possession, distinct self-identity, authenticity, self-legislation. Obviously these would need explication to make clear what is being talked about if we wanted to see how closely they resembled actual self-government and how far they differed. But I think they make the drift of Feinberg's discussion clear.

The third idea of autonomy Feinberg discusses is autonomy as an ideal of character. An autonomous agent is admirable because he has fashioned his own character. But it is notable that a character is not admirable only because an agent has shaped it consistently with his own principles. A self-made character can be a bad one, he can be selfish, cold and ruthless. So the admirability does not only depend on the fact that it is self-made. It depends also on the fact that it is good or noble. An agent who is a cold blooded murderer can claim that he has chosen his principles and to be to that extent autonomous. But he is in no way admirable. Autonomy does not exclude the possibility of an agent living in a society with others. This seems actually to be a minimum condition on autonomy for it to be admirable. It just does not seem to be possible for autonomous agents to be completely independent of each other and to flourish. But this does not in any way detract from the value of autonomy. Feinberg admits all these complications for this concept of autonomy but it has an important role to play.

The fourth concept of autonomy is a right to self-determination, a right to sovereignty. This is conceived

along the lines of national sovereignty in which a state controls a particular area where people live and no other state has a right to interfere. The idea is that a person is sovereign in the same sense as a state is sovereign. It is not quite clear what sovereignty in a state amounts to but Feinberg characterizes is thus: "A sovereign state is territory under a kind of unconditional and absolute jurisdiction."[3] One important question about sovereignty is where it comes from. There does not seem to be any obviously correct answer but one candidate is the nation. Then we need to know what a nation is to which there does not seem to be very clear answers. But in practice there are clear instances of nations like the nation of Finns and their state Finland. So we do not need to worry too much about the exact nature of nations. But the most important feature of this concept of autonomy is the absolute and unconditional nature of it.[4] This fourth meaning of autonomy seems to be external autonomy as I described it above.

These four meanings of autonomy are not mutually exclusive and I hope to show that they can be categorized as either internal autonomy or external autonomy. I said above that I will delay discussing the first meaning Feinberg describes until the next part of this chapter. What Feinberg is hoping to do in discussing autonomy is to discover a notion which justifies a ban on paternalistic interferences. He believes the notion of

3. Feinberg, J.: op. cit. p. 49
4. Feinberg discusses this notion of autonomy in detail in chapter 19 in op. cit. pp. 52-97

personal sovereignty supplies a ban on paternalism which goes further than ensuring that an action is voluntary.

The second notion of autonomy was the actual condition of self-governing as I described it above. The central idea was that the agent actually did take his own decisions and shape his life according to his own values. This is dependent on luck to a great extent: serious illness can prevent an agent from achieving his autonomy and so can dire poverty. But given that the agent is able to do as he chooses in the sense that there are no internal or external barriers to his acting on his choice and he does what he chooses, then he is autonomous.

This meaning of autonomy is rather curious, if I understand it correctly. There is no requirement of goodness or nobility built into it. The only requirement for being autonomous is that the agent decides himself what he does and that there are no barriers against him acting on his choice and he actually does what he decides. This category sits uneasily with the distinction I made between internal and external autonomy. It seems that external autonomy and internal autonomy in some minimum sense of ability to see good reasons are conditions for actual autonomy to obtain. But there is no link between autonomy in this second of Feinberg's meanings and the ability to see good reasons and act on them. What is required is only the ability to decide and to act whatever the reasons.

I take it to be obvious that this cannot be all of autonomy and that it is not Feinberg's intention to imply

this. But in so far as this is autonomy, it seems to be compatible with paternalism. If paternalism limits itself to removing barriers to actual autonomy as described above, then it does not violate autonomy. Wearing seatbelts in cars or protective helmets on motorcycles can help us to act on our own choice. So this meaning of autonomy will not do the work Feinberg wants it to.

Autonomy in the third sense as described above was the ideal of autonomy, autonomy as an admirable character. It is notable that Feinberg does not make any extensive attempt to describe or explain autonomy as an ideal. As I noted above autonomy is compatible with serious flaws of character. This just tells us that autonomy cannot be the whole of the admirable characteristics of character; it requires other things along with it. But it seems reasonable to accept that autonomy is one part of an ideal character. But autonomy both needs to be a part of character to the right degree, neither more nor less, and it needs to be connected to the right traits of character.

But what is this ideal of autonomy which is worth aspiring to? Because it is an ideal it is desirable and it would be valuable to bring it about.

> "The autonomous man is the one who, in Rousseau's phrase, "is obedient to a law that he prescribes to himself", whose life has a consistency that derives from a coherent set of beliefs, values, and principles, by which his actions are governed.

> Moreover, these are not supplied to him ready-made as are those of the heteronomous man: they are his because the outcome of a still-continuing process of criticism and re-evaluation."[5]

There are at least three things in this characterization of the autonomous man. The first is that he obeys his own law. The second is that the actions and beliefs of an autonomous man are consistent and coherent. The third is that his principles are his because they are the outcome of a continuous process of criticism.

The autonomous man does not reach his ideal through abstract theoretical deliberation but through experience from which he learns how he responds himself to different situations, needs and ideas. Even though the ideal is independent of any particular agent, every agent must find his own way of reaching the ideal. There are no fixed paths towards it which he can find and then automatically follow. But neither should we say that the autonomous agent must take a decision before every action however small and insignificant like an existentialist gone mad. It is both that an autonomous agent can be a member of a continuous tradition and accept its standards, even unreflectively, given only that he is ready and willing to examine critically the tradition if and when the occasion arises, and also that he can

---

5. Benn, S. I. : "Freedom, Autonomy and The Concept of Person", in Proceedings of the Aristotelian Society, Vol. lxxvi, 1975/76, p. 124. See also S. I. Benn's book A Theory of Freedom, Cambridge, Cambridge University Press, 1988, chapters 9, 10 and 11, pp. 170-212

internalize his principles. So he has no need to reflect and deliberate before every single decision; successfully internalized principles can guide his actions without prior reflection. Nor does the autonomous agent pick and choose his principles like we choose ties and socks to buy. Principles are not of the same type as socks and ties. We come by them through the learning and experiencing of our own tradition in thinking about the world around us, about ourselves and our society. They mould our beliefs and our thoughts but it is possible to discover these principles and to think about them critically. We sometimes discover that our principles according to which we have lived are incoherent. These discoveries can be difficult and painful, but the response to such discoveries cannot be fully determined by the norms accepted by the society. In such circumstances the agent must reflect and deliberate on his action and be ready to come to a conclusion independently of the accepted traditions in his society, if that is what he thinks is the correct conclusion.

The autonomous agent is distinguished from the heteronomous agent. The heteronomous agent obviously can experience the same kind of difficulties as the autonomous agent, but his distinguishing feature is that he is always looking to the group he belongs to for guidance in finding his answers rather than trying to formulate some himself according to those criteria he is willing to accept.

The three main characteristics of the autonomous agent should be a little clearer by now. His beliefs, principles and values should be consistent or he ought to attempt to make them consistent. The major way towards the end of consistent and coherent beliefs is the continuing process of criticism and re-evaluation. This seems to be the best means to secure this goal. The third part of the ideal of autonomy was that our law should be our own. In what sense should our law be our own? This sense cannot be that everyone must make or create his own law. This would be moral anarchy and I do not think that is a reasonable option. Anyway it is clear that Benn, the author quoted above, is not advocating moral anarchy because he is quite clear that his autonomous moral agent should be a member of a tradition and traditions do not allow the possibility that everyone has their own law different from anyone else's. The sense in which the law or principle is his own is that the agent accepts after rational reflection the law or principles he has discovered in his own moral tradition or outside it as a valuable part of his moral universe.

It is no part of this sense that the agent must have a particular starting point from which he can view all possible moral principles. It is not possible for a moral agent to reflect on or revise all his moral principles at once. To be able to revise and reflect on moral principles he must have some already. He can reflect on a principle when his experience has shown him that there

might be incoherence among the principles he adheres to and follows in his practice.

I do not want to claim that I have answered all relevant questions about the distinction between an autonomous and heteronomous agent or about the relation between an autonomous agent and his tradition. The problems center around the obedience to one's own law and what that comes to. Another question is whether the distinction between autonomy and heteronomy is completely clear. But whatever the answer to these questions turns out to be, it is clear that the autonomy in question is internal autonomy.

Does this account of autonomy exclude reliance on authority? It seems to exclude some types of unreflective reliance on authority. But what about reliance which says that something is right because A says so and A has good reason to know? Now on one understanding the account above would exclude this because it would be required that the agent himself had to know either directly or through some reasoning that this something was right. But I do not think this is a reasonable requirement on autonomy as such or that this is the only understanding of autonomy as described. The reasons for this will become clear. But the principles the agent abides by are never all rationally reflected on. Some principles are, therefore, accepted by relying on authority, the authority of our own tradition or by relying on a religious or moral authority. This seems to me to require that autonomy and reliance on authority be compatible.

Autonomy in this sense forbids all things that prevent critical reflection on moral principles. Political tyranny, for example, would be unjustified. We should notice, though, that political tyranny only goes against the acting on such critical reflection and being free from interference or if such reflection were made public. Political tyranny need not be against individual cases of critical reflection itself in so far as no one else is made aware of this critical reflection apart from the agent. But political tyranny would be against any culture of critical reflection and in so far as the fostering of such a culture is necessary for critical reflection to flourish then political tyranny is unjustified in terms of this conception of autonomy. The two parts of autonomy, internal and external, are both necessary. Having one without the other is not having autonomy in the full sense. Political tyranny, therefore, cannot be compatible with autonomy in the full sense of the term, as I explained it earlier.

This leaves only the last conception of autonomy, autonomy as a right to sovereignty. This notion is taken from the political vocabulary about states and nations as indicated above. The most important feature of it is that a state has an absolute and unconditional jurisdiction over its own territory. When this is transferred to persons it comes down to unconditional sovereignty over the personal or private domain. This means that I alone have the right to take those decisions that concern the private domain only. This is essentially external

autonomy. This does not rule out that I may under some conceivable circumstances stop taking such decisions. I may be overpowered by a tyrant. But it does rule out that this can be done justifiably, unless I voluntarily decide to give up this right myself. This possibility is not problematic neither when this relinquishment is temporary nor when it is permanent. If it is permanent it covers cases like voluntary slavery. This might seem problematic on similar assumptions to Mill's that "The principle of freedom cannot require that he should be free not to be free"[6] if it be allowed that this absolute ban by Mill is on the state enforcement of such contracts rather than on voluntary slavery. But this principle of Mill's seems problematic. And if we take the analogy with states seriously then this possibility does not seem to be incoherent. A state, like Newfoundland let us say, can decide in the appropriate way to become a part of another state like Canada. In doing that the state forfeits its sovereignty but it does not appear to be incoherent or problematic for any conceptual reasons to do so. But this seems to be implied by Mill's objection to voluntary slavery interpreted as above. This possibility of permanently giving up sovereignty does not seem to pose any problems for this conception of autonomy.

As already mentioned this is Feinberg's favoured meaning of autonomy. Why should we prefer this conception of autonomy to the others discussed before? For one thing

---

6. Mill, J. S. : On Liberty, in Three Essays, Oxford, Oxford University Press, 1975, p. 126

it seems to be stronger and to block most, maybe all, paternalistic interferences. But is that an advantage? I am not sure. There seem to be three main reasons for accepting this kind of autonomy as argued by Feinberg.[7] First, it accords well with some of the moral idioms we use: this is no one else's business, I live my life as I please, this is entirely up to me. Attitudes like these seem to be well entrenched in modern consciousness of our status and relation to others. Second, if an agent seriously believes that an action is in his own interest and someone else disagrees and is right, the agent can either be mistaken about a matter of fact or about the nature of his self-interest and its reasonableness. If the mistake is of the first type, then he can be put right and the agent himself ought to welcome it. If it is of the second type, the matter is more difficult. But it is certainly not insurmountable to find out if someone's view of his own interest is rational or not or if what he thinks is reasonable really is reasonable or not. A man who believes that his own best interests are to be able to smoke as many cigarettes a day as he possibly can, has a curious view of his own interests. Even though we accept that few people welcome the overriding of their desires, there is no way to infer from that that all values are equally reasonable or rational. This would simply be false. The third reason for this conception of autonomy is that agents must be allowed to make mistakes. It is necessary for them to be able to make mistakes:

7. Feinberg, J. : op. cit. p. 61-62

"There must be a right to err, to be mistaken, to decide foolishly, to take big risks, if there is to be any meaningful self-rule;..."[8] The idea is that the right to err is a necessary part of the right to sovereignty.

This conception of autonomy seems generally to block paternalistic interferences and if it proves to be justified it will do the work that the other conceptions failed to do. Further work on this notion of autonomy would include attempting to demarcate the private domain and it seems that those who want to use this notion would like it to exclude the obligation to wear seatbelts or protective helmets among other things. This seems a reasonable conclusion in the light of this autonomy. It would also make a ban on the selling of narcotics problematic and a ban on brainwashing. If the agent really wants to inject narcotics into his body and this does not affect adversely anyone else's interest and the agent knows what he is doing, then it really ought to be up to him. The same goes for brainwashing. There ought to be no stopping an agent wanting to undergo brainwashing to do so, if he knows what he is doing. But this kind of autonomy is limited to those who are adult and of sound mind. So it does not stop making children going through education because they are children. But it would stop a well-meaning paternalist from forcing an adult to acquire some education. This type of autonomy seems to get the results that many people believe are obvious truths about ourselves and our private domain.

---

8. Feinberg, J. : op. cit. p. 62

I do not think the three reasons for this notion of autonomy amount to very much. It is certainly true that it accords well with certain idioms of our moral vocabulary. The moral attitudes these idioms express are widespread nowadays. But this does not mean that they are well founded. It is certainly not an obvious truth to be gleaned from our use of moral language, that people ought to be left to do themselves to death if they so desire, important though moral language is. An argument is required to make this thought plausible. The second reason was that when judgements of self-interest depended on the views of the agent and his values, then everything became more difficult than when we were dealing with matters of fact. As I pointed out above, I do not think this will prove insurmountable. There is a limit to what can be counted as reasonable and this puts a limit on what we are willing to accept as rational or reasonable self-interest. I do not want to claim that this clears away all resistance to unlimited paternalistic interferences. It does not. But it throws some doubt on the claim that somehow values cannot be mistaken in a similar fashion to matters of fact.

The third argument for autonomy as sovereignty was that there was a right to err that was a part of the right to sovereignty. I think it should be accepted that the right to err is a part of the right to sovereignty. It is an important part of autonomy to have the opportunity to make mistakes. It is in most cases a good way for the agent to find out whether a certain course of

action is right or wrong. But the problem with this right as formulated by Feinberg is that it does not matter at all how serious the mistake is. It seems to me that it is reasonable to guard against some serious mistakes by ruling them out on the ground that they are against the agent's interests. This seems to be a reasonable policy in terms of self-paternalism and also for paternalism proper.

The last reason against accepting the analogy to national sovereignty as a good way of conceiving of autonomy is that it does not cover the whole territory. I said before that autonomy can be divided into external and internal autonomy. I have not said anything yet about how these two parts of autonomy are related. But it seems a reasonable requirement on any attempt to explain autonomy that it cover all the things we would expect to be covered. The notion of sovereignty as described here only covers external autonomy. It protects the personal domain enabling agents to act on their rational decisions.

It seems possible to explain autonomy as sovereignty in terms of presumptions and aversions. This would obviously not amount to an absolute and unconditional right but in practice in most cases it would be substantially the same thing. Probably the best argument for a presumption against interference is Mill's moral muscles argument, as Feinberg calls it.[9] Mill puts the basic idea thus: "The mental and moral, like the muscular

9. Feinberg, J. : op. cit. p. 384, a note to p. 58

powers, are improved only by being used."[10] To be able to use our moral powers we must make choices and act on them which in turn requires a domain in which to act. This domain is not only to act on rational and reasonable choices but also to make mistakes. Interference with our choices hampers our development of our distinctive talents. Hence, paternalistic interference should be ruled out. This argument makes sense of autonomy in both the senses I have mentioned, the ability to discern good reasons and to act on them and non-interference in the private domain. This ability is not something innate which is operative in every individual irrespective of his surroundings. It only comes to be a discriminative power when properly applied in various contexts and allowed to be acted on. This requires a private domain in which the individual can apply his talents and act on his decisions without hindrance.

This account of autonomy seems to be acceptable unless we find some reasons against it. It seems to me that there are at least two reasons for doubting it.

The first is that autonomy on this argument does not amount to an absolute and unconditional right to do whatever the agent likes. As should be clear by now I do not think that interferences on the ground of the agent's interests or for his own good are ruled out although we can accept a presumption against them. The moral muscles argument by itself does not guarantee an unconditional right, only a presumption against interference.

---

10. Mill, J. S. : On Liberty in op. cit. p. 72

The second is that autonomy understood as an unconditional right to sovereignty in the personal domain has the consequence that when the right and the good of the agent come into conflict the right always has priority over the good. This does not mean that we must believe that the right of the agent and his good are constantly in conflict; only that when they do the right has priority. Because of the nature of the right the option is not open to us to weigh the harm the agent is likely to do to himself against the right and the probable resentment the interference might cause. But the moral muscles argument does not rule this possibility out. It tells us that the harm must be severe to justify a paternalistic interference.

My conclusion to this first part of this chapter is that there is reason to doubt Feinberg's account of autonomy. His preferred notion of autonomy as sovereignty has serious drawbacks. His major claim that autonomy rules out paternalistic interference altogether does not seem to be true. In certain instances they seem to be compatible. This has the consequence that we must decide for each case or for each type of cases under consideration whether a paternalistic interference is justified or not. This ultimately comes down to accepting that in general autonomy is the best guarantee for the good of the agent. But in those cases where we have reason to believe that a paternalistic interference might secure his interest better than leaving it all to himself autonomy does not have automatic priority over the good

of the agent. This strategy is all that is needed for what I have been arguing.

6.2. Capacity and Incapacity

I said above that Feinberg distinguished four meanings of autonomy. One of them I have not discussed. It is autonomy as capacity. I have already indicated that there is reason to doubt Feinberg's account of autonomy but I think it can be profitable to take a look at autonomy as capacity to gain a clearer understanding of the nature of autonomy.

The capacity in question is a natural ability to understand, to have beliefs about ends and means and other things necessary for a normal life. This is the capacity and it obviously depends on the understanding of the capacity how we understand incapacity. Those who are excluded from this category of autonomy as capacity are those who are mentally ill, those who are mentally subnormal, patients who live in a vegetable coma, infants, seniles and generally those who are permanently incapacitated.

Incapacity is caused by impaired or undeveloped faculties. Because the faculties are undeveloped it is not possible for the agent to appraise the features of his circumstances that are salient to the desirability of his action. A young child cannot appreciate the dangers in its surroundings and if left to its own devices, it would sooner or later come to grief. That is why adults

take it to be their obligation to guide the child paternalistically avoiding the dangers in the surroundings and on the way teaching it what to avoid and what to look for and desire. Even though a child cannot articulate it learns as time passes what is in its best interests.

The question is how should we understand this capacity? There seem to be two concepts at work in our understanding. The first is a capacity allowing degrees like IQ. There are those at the lower end like those who are severely retarded and are like children and do not for example understand dangers around them or promises or obligations. Next above them are the mildly retarded who have better understanding of the harm that might befall them, if they crossed the street, for example, or cut a slice of bread. Then we could go up the scale until we have those at the top who are super-intelligent, have the best capacity to discover the important things in the world around them and make those choices that best fulfill their desires. The other concept of capacity at work is an either-or concept. That concept works like legal competence: you either are legally competent to marry a couple, for example, or you are not. There is no other possibility. You either are legally competent to vote in an election or you are not. This concept is at work when we say that most people are entitled to enter contracts, choose a career or do whatever else they think fit with their own lives.

There are clear implications for paternalism flowing from either of these concepts. If we accept the first one it seems to follow that those at the top end of the capacity scale ought to decide for those at lower levels what is in their own best interests because they are obviously the best judges. Incapacity would be relative to the level of capacity of those above you. Large scale paternalism only limited by what would be practicable would follow.

The second concept of capacity is different. It clearly has no obvious unacceptable paternalistic implications. One might think it came to something like the view of autonomy described earlier: autonomy as sovereignty. It certainly seems to block paternalistic interventions because if an agent has the capacity in question we are not apparently allowed to interfere paternalistically. If he has not, then such an interference seems to be in order. But this would be too swift and not, I think, quite correct as I hope will become clear. The threshold concept of capacity, as we can call it, or a nonrelative kind of mental impairment, as Daniel Wikler calls it,[11] does not do all the work that sovereignty does. But to see this we must look a bit closer at what the threshold concept amounts to.

There seem to be tasks that face most human beings in their lives that are such that they require only a

11. Wikler, D. : "Paternalism And The Mildly Retarded" in Sartorius, R. (ed.): Paternalism, Minneapolis, University of Minnesota Press, 1983, p. 87. In this article Wikler formulates the distinction between these two concepts of capacity which I use.

certain amount of intelligence and every increase above that level does not put the agent concerned in a better position to perform that task. So anyone who falls below this threshold and has difficulties in performing those tasks will need protection. The type of task in question is like a decision to marry, if we can call that a task, or decisions to invest on the stock market. It may be true for both of these tasks that above average intelligence or mental capacity will secure a better partner or profits for the one who has such intelligence. But the ability to avoid risk is much simpler and does seem to require only a certain level of understanding of the danger involved in a contract. Dangers in marriage are much more subtle and complicated than in financial dealings, but I think it is not absurd to suggest that they have this feature in common with the financial dealings.

It is important to realize that this threshold is not arbitrary. It is not chosen by me or anyone else but seems to develop in a society and be fixed by it. Even though it is not arbitrary it is conventional and it is certainly true to say that it might be fixed differently from the place where it is fixed now. But this is no easy or simple thing to do. The level is fixed for those of average intelligence so that they can deal with those problems that most people have to deal with in their lives. Whether it is filling out a form for the Inland Revenue or driving a car or buying groceries none of these tasks would be better performed by people of

superior intelligence than those of average intelligence. But for those below average, and average here is understood not as a point but a range, all these tasks could pose insurmountable problems. The reason that the level falls in this place is utility. This is the place where it is most useful for most people to have the threshold. It might certainly fall in a different place but that would involve costs that are unreasonable. If it were lower, some of the institutions relying on the threshold would lose their utility to normal rational agents. If it were higher, a large part of the population would have to rely on opinion and advice from those who understood the tasks and were able to perform them. Many people would be put in the present position of the mentally retarded. This means also that the social practices and institutions in question would lose their utility.

The tasks confronting any normal rational agent in his life in modern society are various and they are of different types. It is often remarked that modern society has become more reflective than societies have been before. This presumably means that it is no longer sufficient to follow accepted practices and traditions unreflectively to live a good life in a modern society. It has been an integral part of the Western intellectual tradition to encourage rational, critical reflection. There are other different traditions where critical thinking is discouraged. It is not traditions as such that somehow demand unreflective consent, acquiescence or

conformity. If a tradition is of the type that requires unreflective conformity, then this will affect those who are brought up and educated within it. But this does not seem to be true of the Western tradition and present Western society, although it may be true of societies of other cultures and Western society in earlier times. This increased reflectivity requires more knowledge of various things than before and it requires more reflection on one's decisions and actions and on what is right and wrong, good and bad.

I do not want to argue whether this is a bad thing or good thing. But if it is true it seems to have one consequence. More knowledge and more thought are apparently required than before. I do not think that man's capacity for thought has increased or become better, but many things have happened which make it easier to assess claims to knowledge and good thinking than before. The instruments we have for evaluating knowledge and truth have become more sophisticated but this in turn makes more demands on our capacity. If our capacity for thinking and coming to correct decisions has not increased, then it seems plausible to say that it is more difficult now than before to reflect and decide. On many important issues we seem to be in a similar situation to American senators who do not have the time to acquaint themselves thoroughly with every issue they have to vote on and sometimes have no idea which way to vote. What they do is to follow the lead of those they trust and vote like them. On many important issues we

seem to be doomed to a fate of following the lead of those in the know. Many people cannot be bothered to think about surrogacy, for example, and they are perfectly content to leave the issues to the decision of the experts. It does seem that no other option is open to anyone nowadays than accept the expert opinions of others about some issues because it is not possible to be well acquainted with every issue one must decide even within the personal domain.

One obvious objection to all this is that it has no relevance to paternalism because paternalism only concerns those things which are directly relevant to ourselves and our good. We are surely able to be masters of our own lives and our own fate in so far as it concerns ourselves. This ability under normal circumstances should ensure that we are not in the position of the American senator. We do know our best interest.

But this is exactly what does not seem to be the case all of the time. Even within the private or personal domain we do not seem to be in full possession of all the relevant knowledge and not to be in the best position to judge our own interest some of the time. Let us take the example of buying our food. Some of the food we buy we do not have the faintest idea what contains. It just happens that we ourselves or someone else in the family likes it and therefore we buy it. Satisfying our desires in the best interest of ourselves depends on a lot of people doing their job properly and informing us about what they

do. Consumer laws often oblige producers to supply information on the packets of the product on the ingredients and the processes used in making it. They also forbid certain practices. If everyone were the best judge of his own interest, no such interference would be justified nor would it be justified to oblige the producer to supply the information about the ingredients. But it certainly is justified to require him to do so and to ban dangerous substances in our food. This is plausible on plain utilitarian grounds. It saves us time not having to check everything and to be able to rely on at least a reasonable degree of non-dangerousness. If the buyer does not want to know, he can avoid reading the information. If he wants to have the option of buying anything he likes, this is not open to him. We do not think it is justified for him to have the option of buying anything he wants.

One point often made in this context is that banning dangerous substances or requiring producers to inform the buyer can be justified on the grounds of harm to others. In selling something you are potentially harming another person. So even though it is not justified to limit the right to buy anything one wants, it is perfectly justified to ban selling anything one wants to sell. Selling is in the public domain and subject to the constraint of harm to others. This seems to me to be true. It comes to the same thing as the argument from utility mentioned above. So even though a paternalistic

argument is possible, there is another argument both possible and plausible.

There is a feature of freedom which ought to be considered in this context because it may throw some light on this notion of autonomy as capacity. Freedom is many faceted and you can be free in one respect but unfree in another. It is not simply a matter of either being free or unfree. One can have economic freedom, complete or limited. One can have social freedom and one can have political freedom. These different kinds of freedom seem to be conceptually independent of each other. Some people have argued that economic freedom is a condition of all other freedom. But this seems to be false. There is at least nothing contradictory about having less or little economic freedom but full political freedom. I do not want to hold, that such a state of affairs would actually be very stable in a society but there does not seem to be anything incoherent about it. I do not need to deny that different freedoms are interrelated. But it is certainly possible to have political freedom and yet at the same time have a highly regulated economy. It may be difficult politically to have a fully planned economy without political freedom being in danger. But the danger is not because of any conceptual incoherence about such a state of affairs.

Autonomy in the sense of capacity seems to be similar to freedom in this respect. It is certainly possible to be autonomous in one area of activity but not in another. The reason is that autonomy in this sense

does depend on natural ability and it is a fact that natural ability is not equally distributed and multi-faceted. So one and the same person can be autonomous in one area and not in another depending on his abilities. There is nothing problematic about saying that A has a natural ability to play football but no gift for foreign languages. But we might reply that it was not natural ability as such which was in question but natural ability to come to reasonable conclusions about one's own interest and shape one's own life. But this more limited ability is still multifarious enough to allow a distinction like the one I made above between different kinds of freedom. This seems to imply that we can have a capacity to decide on one type of tasks concerning ourselves but not on another. If autonomy is natural ability then the same ought to apply.

If this were true of autonomy it would go against well entrenched views of autonomy that have been put forward. I shall discuss the nature of autonomy in the next section of the chapter. But one thing should be made clear at this stage. These two notions of ability or capacity that I have described do not necessarily exclude each other. If we accepted the threshold conception, then we do not need to reject the graded ability conceptions. It is better to conceive of the threshold as a point on the scale of the graded ability. If a person reached that point then it is supposed to be capable of deciding on his own what to do with his life, but if he were below the point then he would need protection and assistance.

As before this point would be socially fixed. There does not seem to be anything incoherent in treating these two conceptions in this way. These two conceptions do not seem to exclude each other.

6.3. The Nature and Value of Autonomy

I think it should be obvious by now that the notion of autonomy is not simple nor does it seem to be very clear. I suggested earlier on that it could be divided into two parts: internal and external autonomy. The discussion of Feinberg's theory has not led to any conclusion about autonomy. His four meanings of autonomy have not forced any theory of autonomy on us. It is high time to have a closer look at the value and nature of autonomy.

I have drawn a distinction between a person's interest and his wants or desires. The interest is objective and it can be damaged without the person's being aware of it and the interest is there whether you attend to it or not. A person's interest is not necessarily unconscious but it can be. Wants or desires are different in that they are in general conscious and they usually do not change or are frustrated without the person realizing this. I do not propose to discuss in detail how this distinction should be drawn. If my money is stolen, then my interest is harmed. This can either happen without my knowledge or not. My desire for a glass of water can be thwarted by the tap turning dry but

generally I would be aware of a thwarting of my thirst. I know the distinction is contentious but I hope it is fairly clear in outline.

On which side of this distinction would autonomy lie? It seems that autonomy should be a part of each person's interests or good. The reason is that autonomy can be violated without the person himself knowing or realizing that his autonomy was being violated. I mentioned earlier a patient who had expressly forbidden the injection of blood into his body even though his life was in danger. The doctor ignored his wishes and injected blood into his body when he had become unconscious and thereby saved his life. In doing so the doctor was violating the patient's autonomy. It seems clear that autonomy ought to be put on the objective side of the distinction between wants and interests. So autonomy is part of the objective welfare of rational agents.

As is clear from the example of the patient and the doctor autonomy can be contrary to other parts of the welfare of agents. Violating the patient's autonomy in this case saved his life and saving your life is probably the clearest instance one can get of something being in one's interest. But it is not only that autonomy can be contrary to an agent's interests. Other parts of an agent's interests seem necessarily to be contrary to autonomy. It can be argued that dependency is a part of an agent's interests and in some instances it can be one of the most important parts of an agent's interests. Dependency, it can be argued, is in direct contradiction

to autonomy. If the agent's welfare includes two parts that are in direct contradiction to each other, then it might be advisable to revise our notion of welfare.

It seems to me that the notion of autonomy involved in the contradiction with dependency and the notion being violated by the doctor's decision do not necessarily coincide. To see that we need to have a closer look at autonomy.

I have already said that autonomy comes in two parts: external autonomy and internal autonomy. Internal autonomy is the ability to see good reasons and to act on them.[12] To be externally autonomous is being free to pursue your projects in the personal domain. This characterisation of autonomy is not obviously true and needs argument. I shall try to give some reasons for considering it in this way.

What is 'the ability to see good reasons and act on them'? The first thing to be said about this is that the analogy to perceptual abilities should not be construed too literally. In 2.1. I rejected the suggestion that our ability to discern moral qualities was essentially perceptual. The same considerations seem to apply to this interpretation of autonomy. Disregarding the literal interpretation of 'see' I do not think there is any objection to use the word. The other important part of the description of internal autonomy was: 'good reasons'. What constitutes good reasons? There can be good reasons

---

12. This formulation of the distinction was suggested to me by John Skorupski.

for beliefs as well as actions. As I have described the internal autonomy the emphasis should be on good reasons for action. But I do not want to indicate that there is some fundamental difference between good reasons for actions and good reasons for belief. As far as I can see there is not.

Good reasons for actions are reasons that have good consequences. The consequences that are good are those consequences that affect sentient beings beneficially. In the case of rational beings with complex value structures it is not easy to describe what these good consequences are. But a good reason would be a reason affecting important values favourably, increasing happiness or autonomy to name two important examples. Further characterisation of good reasons is not necessary in this context, I believe.

But this ability to see good reasons is not a natural ability that every man is endowed with fully fledged, so to speak. To have this ability is an achievement like knowledge or expertise is an achievement. It is clear that all rational beings do not have this ability at the same level or the same amount of it. All people are not interested in good reasons and the natural ability that the achievement depends on is not equally distributed any more than other natural abilities. So we would expect some people to be better at seeing good reasons than others.

But if we are not to take the perceptual analogy too seriously, how should we think of this ability which

constitutes internal autonomy? I think the ability should be thought of as the capacity to understand and use certain moral and evaluative concepts to think critically about reasons and conduct. This critical thinking enables us to make our beliefs coherent, if we are successful, and coherent beliefs make up our claims to moral knowledge. The ideal is to have a coherent set of beliefs in reflective equilibrium. But this is an ideal which is difficult and probably never fully realized by finite beings.

One idea in autonomy is that our laws are our own as I mentioned before. It is not only that laws become our own when we think about them. They become our own when we act on them as well. Acting on our beliefs and principles is probably primary for beliefs becoming our own in the sense that our development of critical thinking comes later than action and enables us to evaluate our beliefs. These two parts of internal autonomy, i.e. the ability to see good reasons and to act, are both necessary for someone to be autonomous.

It should not be inferred from this that a well developed critical faculty is the only necessary thing for achieving autonomy. A certain minimum is certainly necessary. But kindness of heart and a benevolent disposition towards other people would enable most people to see good reasons and act on them in any ordinary life. Critical reflection is not the only relevant factor enabling us to see good reasons. Knowledge is another and kindness of heart a third.

What is the relation between the two parts of autonomy? As I described above internal autonomy comes in grades but external autonomy does not seem to be gradual. If internal autonomy is below a certain level external autonomy does not become operative. Those creatures at the subautonomous level require protection and paternalistic guidance. This applies to children, mentally retarded, seniles and also to animals, it seems. The main reason for this seems to be that they are unable to express and articulate their wants and thoughts coherently, if at all, and some of them are not able to lead their lives coherently. But at some point, perhaps socially fixed, external autonomy is triggered. In general it seems that the more you have of internal autonomy the more valuable external autonomy becomes. But it does not seem to be true that the more you have of internal autonomy the more you have of external autonomy.

It is possible to argue that the more internal autonomy one has the more external autonomy follows. Then we assume that external autonomy is identical with liberty or opportunities to do what we want. It seems to make sense to say this about external autonomy in the sense of liberty. Increased internal autonomy might enable us to spot more opportunities than we would otherwise do. "But autonomy cannot be identical to liberty for, when we deceive a patient, we are also interfering with her autonomy. Deception is not a way of

restricting liberty."[13] But this is only plausible for internal autonomy, not for external autonomy which is negative freedom. A person who is put into prison, not being told that one of the doors out is unlocked, is free to go but he is not able to do what he wants because he does not know. Deception and manipulation are ways of interfering with internal autonomy but not with liberty. They do not affect external autonomy.

We should notice one consequence of formulating autonomy in the way I have done. Let us look back at the example of the Jehova's Witness who did not want any blood injected into his body. The doctor's action is clearly a violation of his autonomy. But because it is the ability to see good reasons, not just any reasons, autonomy is not just "the ability of patients to decide on courses of treatment"[14] on any grounds whatsoever. If they are plainly irrational grounds as in this example, the patient is externally autonomous and he has internal autonomy but it is limited by this irrational belief. Internal autonomy seems to require rationality.

Now it is high time to look at the two examples I mentioned above of autonomy and its status. One was of the doctor and his patient who did not want blood. The other was of autonomy being a part of objective interests that included other parts that seemed to be contrary to it. I do not think it can be doubted on the conception I have described here that the doctor violated the

13. Dworkin, G.: The Theory and Practice of Autonomy, Cambridge, Cambridge University Press, 1988, p. 14
14. Dworkin, G. : op. cit. p. 14

patient's autonomy because it seems to be true that he manipulated the patient. But dependency does not seem to me to clash with autonomy as I have described it. If the agent's best beliefs tell him that he should be dependent on somebody else in some parts of his life or that he has to, then I cannot see how this can be contrary to his autonomy. It is only if we think that self-sufficiency amounts to autonomy, that these two things become contrary to each other. But this is no part of my conception of autonomy.

Dependency can come about in different ways. Some older husbands are so dependent on their wives that they cannot pour themselves a cup of coffee without assistance. And some modern wives are so dependent on their husbands that they cannot cook a proper meal. This may not be a result of any conscious choice but has come about in the course of the marriage. The husband or the wife who is relied on may take this as a way of expressing his/her love rather than as an oppressing chore. The one who is dependent can make this up in other fields of knowledge or skill.

A dependency of a different sort develops when someone is an invalid or breaks down mentally in depression. Depression can obviously damage internal autonomy, make you unable to recognize good reasons and act on them, and create a need for protection. In this case there is a clash between autonomy and dependence. But I do not think this causes any problem for what I have said about autonomy.

There are other values which might be taken to be contrary to autonomy. One is loyalty, another is reliance on authority. One might even think that obeying the law is problematic for an autonomous person. These claims seem to be false. Loyalty can be a desirable and rational bond between people. Loyalty is trust and support of one person for another or one sentient being to another. It does not seem to be exclusive to persons. It is no part of loyalty to deny any faults or wrongdoings of the one you are loyal to but you feel that you owe him support and trust, even love. Loyalty is not like an overwhelming desire which you cannot resist. If it were, then loyalty would be contrary to autonomy. Nor is it blind.

Obeying the law does not clash with autonomy. If we took autonomy to require explicit consent to anything we obeyed, then it would be heteronomous to obey the law. But it is not a reasonable requirement on anyone explicitly to have to consent to every law for it to be valid any more than it is reasonable to require a decision before every action of an agent. Laws require legitimacy and it is also true that every agent can break the law if he wishes on pain of punishment. So obeying the law does not contradict the notion of control which is central in autonomy nor the notion of knowledge.

But what if the idea is not about positive law but moral law and somehow the agent had to consent to every moral law to be self-legislative? "...the argument for the necessity of consent is based on the idea of natural rights; in other words, it presupposes an objective moral

law. It cannot therefore be used to support an assertion of moral self-legislation, which is in effect an assertion of the subjectivity of moral law."[15]

But what about the relation between the autonomous agent and authority. This is often supposed to be problematic. The claim is that the autonomy of moral judgement can be expressed by saying that "the moral agent has a special freedom in the area of moral thinking."[16] Hence no moral authority is possible. Another version of the same claim is to say that to accept moral authority is to give up autonomy.

There are various possible arguments for the claim that in moral thinking we have special freedom.[17] But I take it to be clear from what I said earlier in chapters 1 and 2 about objectivity and about knowledge and expertise that this is an argument I want to reject. It seems to me to depend on a false dichotomy between science and morality. It is always possible in science to reject a view of the facts, especially when they are complex. One might even want to reject the fact that the earth is round and decide to rely on some silly hypotheses explaining away the roundness. This would be an instance of the rejection of the norms and canons of scientific rationality. The same is possible in morality. I might accept that I took away your wallet without your consent, that I really stole your wallet but yet reject

---

15. Downie, R. S. and Telfer, E. : "Autonomy" in Philosophy, Vol. XLVI, No. 178, October 1971, p. 298
16. Downie, R. S. and Telfer, E. : op. cit. 295
17. Downie, R. S. and Telfer, E. : op. cit. p. 296-299 where the authors consider three and reject them all.

that stealing was wrong. This would be just as silly as rejecting the roundness of the earth but I am certainly free to reject in both cases.

It seems to me that one formulation of the notion of autonomy contributes to the plausibility of this claim about special freedom in moral matters. The notion of autonomy is often formulated as the fact that in the personal domain it is entirely up to me what I decide to do. This is true, it seems to me, and I have accepted that control of actions and refraining from action is a central part of autonomy. But it seems that this formulation is often taken to imply that the personal domain is value-neutral, that what I do in the personal domain is neither right nor wrong. This seems to me to be false.

I mentioned stealing above. Stealing necessarily involves another agent: you cannot steal from yourself. It may be thought that it applies generally that you cannot do any wrong to yourself. Prudential reasoning reigned supreme in the personal domain without any constraints. It is certainly true that prudential reasoning applies in the personal domain if it applies anywhere. But this does not mean that one cannot do any wrong to oneself. The relation of harming for example applies both in many person cases and in one person cases: it is possible to harm oneself. This becomes clear when we look at killing. There is a moral ban on intentionally killing another person unless circumstances are exceptional like war. But is there a similar ban on

killing yourself? Is killing yourself just as morally blameworthy as killing somebody else? It seems to me that you have some leeway with yourself and your own life which you do not have with other lives and other agents. But this does not mean that moral notions are inapplicable at all in the personal domain. Killing yourself can be a wrong. So the personal domain is not morally neutral. "...private and public morality are therefore not two moralities, but two aspects of a single fundamental moral principle."[18]

One inference from this could be that there is no essential moral difference between the personal sphere and the public sphere. It also seems to be true that there does not seem to be any fundamental difference between science and morality. The status of authority in science is not problematic. Hence the status of authority in morality ought not to be problematic. It seems to be clear that autonomy does not rule out authority. But how should we describe the relation? Is it possible to be related to authority in a way which violates your autonomy?

We can rely on authority in the sense of "x is right because A says so and he has good reason to know." This does not seem to be problematic for autonomy in the sense I am thinking about. This implies that we have checked that A's judgements have proved reliable in the past and that he is likely to handle the task at hand

---

18. Downie, R. S. and Telfer, E. : Respect for Persons, London, George Allen and Unwin, 1969, p. 93

appropriately. It is just rational to accept that others can be more knowledgeable in some, maybe most matters, than we are. This includes morality. It can be a good reason for accepting a belief as well founded that it comes from someone who is in a position to know.

I said earlier on that in a sense acting on good reasons seems to be primary to seeing good reasons. This seems to be one structure in internal autonomy. It seems clear that we have been acting on what we unreflectively take to be good reasons for many years before we start reflecting on these reasons. When we start making our moral views coherent through criticism we are attempting to see whether they really are good reasons. Full internal autonomy always comes late in the day, so to speak, in the sense of rational autonomy. But it is possible to see whether someone is autonomous from what he says and what he is prepared to defend when challenged, even though he has no formulated idea of moral autonomy. Moral autonomy can be latent in habits of acting and thinking.

There is another reason for wanting to accept moral expertise and moral authority. At any stage in time morality can generally be divided into two. One part is formulated in rules and well established principles accepted by those who live by them. The other part is where no principles have been formulated and no well established rules are accepted by all. This latter part covers nowadays things like surrogacy, abortion, nuclear war. It is to be expected that moral experts are more

knowledgeable about morality in general than laymen but in particular they should be able to articulate the problems, arguments and proposed solutions in the area where there are not well established rules and principles. Those who are not interested in such debates might be well advised to accept suggestions from those experts in those debated areas. The status and activity of moral experts as moral authorities is no threat to autonomy.

Knowledge and control over one's own desires and actions are essential to autonomy. In so far as increased knowledge enhances autonomy, moral experts ought to be good friends of moral autonomy. Moral experts articulate moral views and arguments. If they are up to the task they do so in a civilized and reasonable way. The moral experts will be examples to others about what are reasonable moral beliefs. They will also present alternatives to accepted practices which can be considered, reasoned about and accepted or rejected, as the case may be. So the moral experts should help people to see more alternatives and opportunities than they would otherwise have. But at the same time they uphold moral standards. The exploration of alternatives does not mean that all new opportunities will or must be accepted. They might be rejected as inferior. Moral experts can be very important for fostering a culture conducive to rational autonomy.

6.4. Autonomy and Paternalism

In discussing paternalism one must make certain assumptions about those who are subject to paternalistic interventions. First, we can often assume that those subject to paternalism are incapacitated, either temporarily or permanently. This applies to those who are drunk or under the influence of narcotics and those who are mentally retarded, seniles and children. Second, we can assume that the agent is making a mistake about a fact and would welcome being in possession of the truth. Third, we can sometimes assume the agent has an unreasonable view of his own good. If someone made it his life's ambition to stand still on one foot or to smoke as many cigarettes as possible then we would infer that his view of his own good was curious, even unreasonable.

This third assumption introduces the latent elitism in what I have argued. If it is possible to reflect on and evaluate ends just like means then it seems to follow that some values and ends are better than others and some people are better judges of ends and values than others. Now, this means that it is impossible to argue that all people are owed equal respect on the basis of equally good ends and that these ends are equally good because they are not open to deliberation.[19] This theory is false, if my account has been true. The account of

---

19. Skorupski, J. : "Liberal Elitism", photocopy, pp. 1-4 and 27-30

autonomy I have discussed supports this because internal autonomy comes in degrees and as soon as external autonomy comes into play the agent is not to be interfered with. The acceptance of this difference of ends does not mean that we have to give up the notion of the equality of respect. Something like it is behind the notion of external autonomy. But we must find a new basis for it.

I said above that the connection between internal and external autonomy seemed to be that the more you have of internal autonomy the more valuable external autonomy becomes. External autonomy implies that a principle of non-interference applies to the agent and his actions should not be interfered with unless something serious would follow them. Then the weighing of the consequences comes into play and considerations about rationality of the decision. It seems to me that for the simple reason that autonomy is valued highly by most people nowadays the consequences must be very serious for a paternalistic intervention to be justified. If autonomy was generally less valued it would be easier to justify a paternalistic interference.

There is a question about the status of those who are subautonomous so to speak. I take it that in general the principle of non-interference does not apply to them and it would depend on the judgement of someone other than themselves whether they should be guided or not. What relation obtains between their interests and the autonomy of those who reach the threshold is not clear to

me. Would it be permissible to sacrifice the autonomy of one for the interest of five or ten of those who are subautonomous? I do not have an answer to this question.

I mentioned equal respect above. Equal respect can be thought of as a reason for a principle of non-interference or a side constraint. Side constraints put a limit on what can be done to a person and they reflect the separateness of lives and that persons ought always to be treated as ends and not merely as means.[20] These can also be considered as the grounds for autonomy.

Non-interference of moral agents can be thought of in at least two ways. On the one hand it can be thought of as absolute constraints and under practically no circumstances are they to be violated. This is Nozick's view. Nozick's exception from the non-violation principle is moral catastrophes. But otherwise he believes that individuals are inviolable and hence these constraints ought never to be overridden. This is a very strong view. On the other hand it can be argued for on maximizing principles. Then the view is that you ought to minimize the violations of non-interference and maximize autonomy.

I do not propose to go thoroughly into this issue. But I want to mention three reasons for preferring the maximizing view. The first is that sociality is a value and it is a great part of most people's lives. Love and friendship are among the things we treasure most in our lives. An integral part of them is that other people's

---

20. Nozick, R. : Anarchy, State, and Utopia, Oxford, Basil Blackwell, 1974, pp. 29-34

interests become a central part of our lives. These values require us to prefer other persons' projects to our own in some circumstances as an expression of our love or friendship. It is also implausible that we are not required to put aside our own projects to prevent a great harm if it is within our powers to do so. But this seems to be required by the strict understanding of side constraints along Nozick's lines.[21] This is no reason for maximizing autonomy, only a reason against this very strict inviolability.

The second is that in a world without a fully determinate moral code moral conflicts are unavoidable. If a man in an aeroplane with 200 passengers says he has a bomb and that he will blow it unless certain conditions are met and you are in a position to prevent this by killing him, then it seems to be justified for you to kill him. We assume that all the passengers will die if

21. D.O. Brink says in Moral Realism and The Foundation of Ethics, Cambridge, Cambridge University Press, 1989, p. 289, that "Nozick claims that recognition of the separateness of persons implies that no one...can ever be required to set aside her projects in order to benefit others or prevent harm to them." Brink refers to pp. 30-33 of Nozick's book, op. cit. But it is clear from what Nozick says that he is not claiming this. He is claiming that separateness of lives outlaws physical aggression, bans the use of one person by another and that personal sacrifice cannot be justified by reference to a social entity. This is not claiming that one can never be required to set aside one's projects for the benefit of others. The underlying idea in Nozick, it seems to me, is that such setting aside cannot be a moral obligation. It is this idea which seems to imply the radical separateness, a separateness all the way down, that Brink says Nozick claims. Nozick does not claim this idea, at least not in the referred pages, although he allows for moral obligation to others in case of moral catastrophes. But it is this idea which is made implausible by the value of sociality and our obligation to others in hours of great need.

the bomb goes off. It is plainly better that one is killed rather than 200. This seems to show that side constraints do not hold in such a situation and we should take the maximizing view.

The third reason is that it seems to be justified to limit autonomy under certain circumstances to enhance it in the future. I mentioned earlier an example of a ban on duelling. It seems to be true of this that it leads to better chances for the individuals subject to the ban to develop their other interests. It might be objected that the duelling example is not strictly speaking an example of a restriction on autonomy. But it is easy to construct one along similar lines. Let us think of a politician in an election who decides not to tell the public that he is going to raise taxes after the election if he is elected, and he has good reason to believe that the course which his opponents propose to follow will lead to disaster. Let us also suppose that raising taxes will avoid disaster and that his telling the electorate before the election will damage his chances of getting elected. He is elected. After the election he decides to raise taxes. If we agree that these assumptions are true, then he seems to have a justification for not telling the public the truth about his intention beforehand but this also seems to be a betrayal of the autonomy of the electorate. They ought to know the truth before making up their minds. But this might also in the long run minimize the

loss of autonomy.[22] These three reasons seem to lend some plausibility to the claim that we ought to maximize autonomy rather than take the side constraint view of Nozick.

I said that those who could be described as subautonomous are persons to whom the principle of non-intervention is not applicable. It can be argued that it is in their interest to develop their abilities and internal autonomy to reach the level of external autonomy. Autonomy is certainly a key to many of the most important goods of a rational being's life. But what do we say to someone who accepts that he ought to be autonomous but at the same time chooses not to?[23] Can we say that he is wrong?

I think at least two claims can be made about this man. The first that he might be preferring other values like loyalty or dependency of some sort to developing his autonomy or his rational faculties. This man is not losing his autonomy by his choice but preferring a different balance in his life between competing goods to

---

22. I realize that these three points raise many questions that need to be answered. Perhaps the most serious questions surround the notion of agent-neutrality and autonomy and the maximizing view. I propose an answer along similar lines to R. Lindley in his book Autonomy, London, Macmillan, 1986, pp. 82-91 where he accepts interpersonal neutralism but rejects temporal neutralism. There is also more discussion about agent relativity and agent neutrality to be had in S. Scheffler (ed.) Consequentialism and Its Critics, Oxford, Oxford University Press, 1988.

23. Dworkin, G. : op. cit. p. 39 where Dworkin considers this to be a paradox for theories of autonomy that rely on the agent's will or decision. This does not seem to me to be a paradox for a theory similar to the one I have been arguing for.

somebody else who prefers his autonomy. The second is that if he is seriously proposing that he become like a child or a mentally retarded person, then it is not difficult to show that it is in his interest to retain his autonomy. He is a better man if he is autonomous and it is also better for him. He can live a better life. We could also say that he has the duty to be autonomous and develop his talents.

I take it to be true that it is good for a rational agent to be autonomous. This might be taken to mean that Platonic Forms of an ideal person could be discovered and it would be in virtue of those Forms that we truly say that it is better for him. But the notion of objective interests should not be understood to imply this. What it implies is that if a non-autonomous agent were subjected to the experiences of autonomy he would desire it. So objective interests do not totally exclude desires but allow them in at another level than ordinary desire theories.

What should we say about the examples of protective helmets and seatbelts in the light of what the description of autonomy I have given? It seems clear that the obligation to wear both is an intrusion of autonomy because it is not up to us to decide this. The most plausible answer, it seems, is that these decisions are not very important and they do not raise a serious issue for autonomy. So even if we are forced to wear helmets and seatbelts, this is not a great loss of autonomy. Other interests which are possibly enhanced by this

obligation can outweigh that loss. Autonomy does not automatically outweigh other values and interests.

The ban on brainwashing is an instance of paternalism which can be justified in terms of autonomy. This ban can indeed possibly maximize autonomy. The same applies to a ban on manipulation and subliminal advertising. So paternalism and autonomy are not contradictories. Paternalism can enhance and promote autonomy.

6.5. Conclusion

In this chapter I have argued that Feinberg's theory of autonomy and his four meanings of autonomy is not satisfactory. But his theory needs detailed discussion because it touches on most of the important issues concerning autonomy. Autonomy should be divided into two parts: internal and external autonomy. External autonomy is being free from the interference of others. Internal autonomy is the ability to see and act on good reasons. Autonomy does not completely outlaw paternalistic interferences because it is only one value among others and paternalism can be used to maximize autonomy.

Conclusions

In this essay I have attempted to argue that objective consequentialism gives the best account of the good and moral qualities in general. This seems to make moral knowledge and moral expertise possible. Paternalism seems to present some special problems but in general they seem best dealt with in terms of the consequences of paternalistic actions compared to the alternative course of self-determination. The autonomy of persons puts a constraint on any paternalistic action but it does not seem to rule it out altogether. This is the general structure of the essay and its conclusions.

I think it is worthwhile to go over some of the major points again and add some further considerations on paternalism, objectivity and authority.

Some people have been sceptical of objectivity in morality because they believed it led inevitably to intolerance. The idea seems to be that because there might be correct answers to moral problems, then it would be justified to force those right answers on those who disagreed. The government in any state could claim that it was forcing a conception of the good on its subjects because it was the right one. This would be moral authoritarianism and moral tyranny and would obviously be unacceptable to any reasonable person.

> "The fact that many people fear that if they concede any sort of moral objectivity out loud then they will find some government shoving its notion of moral objectivity down their throats is without question one of the reasons why so many people subscribe to amoral subjectivism to which they give no real assent."[1]

Various reactions are possible to this. One is to go for subjectivism in morality accepting that subjectivism implies that there are no right answers in morality. This would certainly exclude any possibility of moral expertise or moral authority. But the point is that on this assumption it would not be possible to argue that it was wrong for a government to force its conception of good on its subjects because it would be meaningless to say it is wrong. There simply is no moral wrongness or rightness. Hence the government could not be wrong. In fact this type of subjectivism seems to leave nothing but naked power to any government or authority to force through some preferred end. So I think that subjectivism is an inappropriate response to this question.

Some subjectivists claim that their theory is the only one which can possibly accept fallibilism about moral views and tolerance. To claim this two things must be true, it seems. First, that mistakes are possible about moral matters. Second, that tolerance is right and

---

1. Putnam, H. : Reason, Truth and History, Cambridge, Cambridge University Press, 1981, p. 149

appropriate but intolerance not. Neither of these things can be true on the subjectivist account. No mistakes are possible about morality and it cannot be true that intolerance is bad or wrong. According to the subjectivist there is no truth of the matter and morality is up to each individual. So he can lay no claim to the virtue of tolerance or the humility of accepting that he might be mistaken about a moral issue.

So subjectivism has little to offer in this respect. But if we accept objectivity in moral affairs, does this have any of the serious consequences that have been claimed to follow? It seems to me that none of the consequences follow that have been believed to follow. First, accepting moral objectivity makes moral mistakes possible. The idea in the objection to objectivity leading to intolerance seems to be that an agent who believes x to be morally right must believe that somebody else who disagrees is wrong. It also seems to be implied in this that the agent ought to be willing in the normal circumstances to act on his belief. It is claimed that intolerance follows from this. But this is false for three reasons. The first is that an agent who accepts that what he believes can be objectively right can also recognize that he could be wrong. Even on the best evidence there is no guarantee that his views are the true ones. The second is that even though it might be true that what A believes is wrong it does not follow that making A do what is right is right. He might resent it so fiercely that there was no value in making him do

what was right. So an agent who accepts objectivity can understand and accept the value of tolerance. A third reason is that on the objective conception of morality rational appreciation of morality becomes possible. So it is always a possibility that an agent who was wrong could be rationally persuaded about what is right and true. It is also true that there is no reason to be dogmatic though you believe you are right. It is a mark of rationality to be willing to subject your beliefs to the test of discussion and persuasion with an open mind.

There is one more reason for any moral realist to be reluctant to use paternalism or to be intolerant about others' conduct. It is that there does not seem to be any unified ordering of the objective goods of rational agents possible. This does not mean that there is a complete anarchy of objective goods. But there is a possible spectrum for agents to choose without crossing the limit into the unreasonable or the irrational. So there is no hierarchy of values which can be used to justify opposition to or oppression of deviance.

It should come as no surprise that moral subjectivists want to avoid a barren theory like the one I described above. I discussed first in this thesis two attempts by prominent contemporary philosophers to use subjectivist assumptions to build up a full blown moral theory. R.M. Hare believes he can use the analysis of moral language as a basis for a moral theory. Hence, the importance of his analysis of 'good' and his distinction between the evaluative and descriptive meaning of moral

terms. I think there are reasons to reject this analysis. This does not mean that one should reject all the other parts of Hare's theory. I am willing to accept that moral principles are universalizable and overriding even though they are not always acted on.

Simon Blackburn argues that he can accept moral language at face value and base it on a subjectivist metaethical theory given the condition that the distinction between first- and second-order theories or ethical and metaethical theories is very sharp and cannot be crossed. No theory can cross the line without damage. He thinks that any second-order theory can go with any first order theory. I think this dividing line cannot be sustained. It also seems true that if Blackburn's theory is successful we would have difficulty believing it. So I think these two theories can be rejected.

Realist theories are of different types. I considered four types of realist theories. Moral qualities have been thought of as analogous to perceptual qualities. This seems to me ultimately to fail. They have been thought of as analogous to primary qualities. I want to reject this analogy although I think moral qualities can play many of the roles primary qualities play. They have also been considered as the entities corresponding to the most reasonable hypothesis explaining agents' conduct. This is something that has to be taken into consideration but this position has its weaknesses. Lastly moral qualities have been believed to be real and particular. This amounts to denying that moral principles

can be general. This seems to me to be wrong. Any moral principle would be such that it allowed generalization.

Moral qualities are objective. There is rightness and wrongness independently of our judgements, knowledge and our minds. Actions and agents are bearers of moral qualities. It can be argued that all sentient beings are bearers of some moral qualities, even that nature itself has some moral qualities. It seems that given some plausible assumptions about rationality, persons and objectivity, we can conclude that consequentialism is the best account of the nature of moral qualities.

In addition value seems best accounted for in objective terms, not in terms of desires. Accompanying this we should distinguish between wants and desires on one hand and interests on the other. Moral values like happiness or autonomy are objective and independent of our desires. These moral values ought to be maximized by any rational, moral agent. Objective values are not Platonic forms discoverable by reason and completely cut off from desire. The objectivity of values should be interpreted as values that would be desired were the agents exposed to them subject to the condition laid down in the introduction.

Because there is truth to be had in moral matters, then moral knowledge becomes possible. Moral knowledge should be accounted for in similar terms as other knowledge. Knowledge is not foundational in the sense that there are some incorrigible beliefs that all other beliefs are based on. It is always possible that we are

mistaken in our beliefs. The claims to knowledge must be evaluated in terms of the coherence of their justification. Although all justification seems ultimately to be circular not all justification is viciously circular. Moral knowledge is subject to the same considerations as other knowledge.

If moral knowledge is possible it seems that moral expertise will follow. Moral expertise has had few followers and most of those who have discussed it have tried to resist it. The most remarkable feature of the arguments against moral expertise is that they do not seem to show any feature of moral knowledge or moral expertise which makes it seriously problematic. The point of these moral experts is that they are the authorities on moral matters. It is not reasonable to expect them to agree on everything any more than other experts. But they enrich moral life by pointing out alternatives and articulating opinions and theories about the part of morality which is new and has not been formulated into principles. But moral experts are not ultimate authorities able to adjudicate on all moral matters. There are no such authorities.

It needs to be pointed out that in this essay I use the notion of authority with the case of expertise primarily in mind. It can be argued that the case of expertise is not central to the notion of authority.[2] Proper authority in a de facto sense exists when an agent

2. Benn, S. I. : "Authority", in Edwards, P. (ed.): The Encyclopedia of Philosophy, New York, Macmillan Publishing, 1967, Vol. 1, p. 216

recognizes that another is entitled to command him. Authority based on knowledge says that you accept a verdict of an expert because he knows what he is talking about. Questions arise for this kind of authority for an autonomous persons but I think it is possible to argue that if in general the acceptance of such an authority has good consequences then it ought to be obeyed. This means that the agent must be prepared at any time to assess the consequences himself and come to a conclusion about the goodness of the consequences. At the minimum it must always be possible for him to do so. So it seems that the same or similar considerations apply to authority proper and to authority as I have discussed it.

Now, if it is true that persons have interests or objective ends and that it is possible to have moral experts who are good at spotting those ends does this not lead to the tyranny of the moral experts? This does not seem to be the case. If we look at the case of paternalism, i.e. making somebody do what is in his own interest, then we have a test case of this possibility. There are two other possible cases that might be looked at. One is that people should be prevented from harming others. I assume that this case is not problematic and it is justifiable to do so. The other is to make them do what is in others' interests rather than preventing harm to them. The crucial difference between this case and paternalism is that what the agents could be made to do would not be in their own interests but in somebody else's. This possibility raises pressing problems about

agent-neutrality of moral goodness which I have only discussed very superficially. So paternalism presents a case where we can check the plausibility of the claim about the tyranny of moral experts.

Paternalism is to be made to act in one's own interest. It does matter how this takes place. There are various possibilities. Coercion is one, coercive threats is another, manipulation is a third. All these would count as paternalistic ways of causing one to work in one's own interest. But rational persuasion would not, if we can take that as an instance of making somebody do what is in his own interest. This is different from the question of understanding whether a paternalistic act is justified. This requires taking the consequences of the proposed intervention and weighing them against the value of the consequences of allowing the agent to decide himself. The rationality of the decisions has weight as well. If in general it can be argued that an intervention has a far higher value than allowing the agent to decide himself what to do the intervention seems to be justified. But the harm from the decision must be severe and it must also be considered that there is always a danger involved in paternalism.

Autonomy is one among the objective goods of any agent. Autonomy includes control of the agent over himself and knowledge by the agent about what concerns him. It is obviously problematic for paternalism whether it is compatible with autonomy. But there seem to be four reasons for accepting them to be compatible. The first is

that autonomy is only one among many various goods of the agent and it is clearly important for a happy and worthwhile life. But it has no automatic priority over all other goods. The second is that the consequences of allowing the agent to proceed on the basis of his decision can be so severe that they outweigh the good of autonomy. Third, any paternalistic action must be in the agent's interest. This means that not any presumed good can be imposed on the agent. It must be in <u>his</u> interest. Four, maximizing autonomy seems to be possible.

Bibliography:

Anscombe, G.E.M.: "On Brute Facts", in Analysis Vol. 18, Nr. 3, 1958, pp.22-25; also in Anscombe G.E.M., 1981.

Anscombe, G.E.M.: "Modern Moral Philosophy" in The Collected Philosophical Papers of G.E.M. Anscombe Vol. III, Oxford, Basil Blackwell, 1981

Archard, D.: "Paternalism Defined", in Analysis, Vol. 50, No. 1, January 1990, pp. 36-42

Arneson, R.J.: "Mill versus Paternalism" in Ethics, Vol. 90, 1980, pp. 470-489

Benn, S.I.: "Authority", in Edwards, P. (ed.): The Encyclopedia of Philosophy, New York, Macmillan Publishing, 1967, Vol. 1, pp. 215-218

Benn, S.I.: "Freedom, Autonomy and the Concept of a Person", in Proceedings of the Aristotelian Society, Vol. LXXVI, 1975-1976, pp. 109-130

Benn, S.I.: A Theory of Freedom, Cambridge, Cambridge University Press, 1988

Benson, J.: "Who is the Autonomous Man" in Philosophy, Vol. 58, 1983, pp. 5-17

Blackburn, S.: "Reply: Rule-Following and Moral Realism" in Holtzman and Leich, 1981, pp. 163-187

Blackburn, S. : Spreading the Word, Oxford, Clarendon Press, 1984

Blackburn, S.: "Errors and the Phenomenology of Value", in Honderich (ed.), 1985, pp. 1-22

Blackburn, S.: "Supervenience Revisited" in Sayre-McCord, G. (1988) pp.59-75; appeared originally in Hacking, I. ed.: Exercises in Analysis: Essays by Students of Casimir Lewis, Cambridge, Cambridge University Press, 1985.

Blackburn, S.: "Attitudes and Content", in Ethics, April 1988, Vol. 98, pp. 501-517

Blanshard, B.: "Wisdom", in Edwards, P. (ed.): The Encyclopedia of Philosophy, Vol. 8, pp. 322-324, New York, Macmillan Publishing, 1967

Bond, E.J.: Reason and Value, Cambridge, Cambridge University Press, 1983

Bond, E.J.: ""Good" and "Good For"": A Reply to Hurka" in Mind, Vol. XCVII, No. 386, April 1988, pp. 279-280

Brennan, J.M.: The Open-Texture of Moral Concepts, London, Macmillan, 1977

Brink, D.O.: Moral Realism and The Foundation of Ethics, Cambridge, Cambridge University Press, 1989

Brock, D.: "Paternalism and Promoting the Good", in R. Sartorius (ed.), 1983, pp. 237-260

Brock, D.: "Paternalism and Autonomy" in Ethics, Vol. 98, April 1988, pp. 550-565

Brown, J.M.: "On Applying Ethics" in Evans J.D.G. (ed.) 1987, pp. 81-93

Buchanan, A.: "Medical Paternalism", in Cohen, M., Nagel, T. and Scanlon, T., 1981, pp.214-234

Burch, R.W.: "Are There Moral Experts?", in Monist, Vol. 58, 1974, pp. 646-658

Callahan, J. C.: "Paternalism and Voluntariness", in Canadian Journal of Philosophy, 1986, Vol. XVI, pp. 199-220

Campbell, R. and Collison, D.: Ending Lives, Oxford, Basil Blackwell, 1988

Carter, R.: "Justifying Paternalism", in Canadian Journal of Philosophy, Vol. VII, 1977, pp. 133-145

Christman, J. (ed.): The Inner Citadel, New York, Oxford University Press, 1989

Clark, S.R.L.: "Abstract Morality, Concrete Cases" in Evans J.D.G. (ed.), 1987, pp. 35-53

Cohen, M., Nagel, T. and Scanlon T.: Medicine and Moral Philosophy, Princeton, Princeton University Press, 1981

Dancy, J.: "On Moral Properties", in Mind 1981, Vol. XC, pp. 367-385

Dancy, J.: Ethical Particularism and Morally Relevant Properties", in Mind 1983, Vol. XCII, pp. 530-547

Dancy, J.: Introduction to Contemporary Epistemology, Oxford, Blackwell, 1985

Dancy, J.: "Two Conceptions of Moral Realism" in The Aristotelian Society, Supplementary Volume, 1986, p. 167-187

Dinan, S.A.: "The Particularity of Moral Knowledge", in Proceedings of The American Catholic Philosophical Association, Vol. 58, pp. 65-72, 1984

Downie, R. and Telfer, E.: Respect for Persons, London, George Allen and Unwin, 1969

Downie, R. and Telfer, E.: "Autonomy" in Philosphy, Vol. XLVI, No. 178, October 1971, pp.293-301

Doyle, D.: "Pain Control" in The Practitioner, 22 March 1990, Vol. 234,

Dworkin, G.: "Paternalism", in Laslett, P. and Fishkin, J. (eds.), 1979, pp. 78-96

Dworkin, G.: The Theory and Practice of Autonomy, Cambridge, Cambridge University Press, 1988

Eggerman, R. W.: "Moral Philosophers and Moral Advisers" in Metaphilosophy, Vol. 10, No. 2, April 1979, pp. 161-168

Evans, J.D.G. (ed.): Moral Philosophy and Contemporary Problems, Cambridge, Cambridge University Press, 1987

Feinberg, J.: The Moral Limits of the Criminal Law: Harm to Self, Vol. 3, New York, Oxford University Press, 1986

Feinberg, J.: The Moral Limits of the Criminal Law: Harmless Wrongdoing, Vol. 4, New York, Oxford University Press, 1990

Foot, P. ed. : Theories of Ethics, Oxford, Oxford University Press, 1967

Foot, P. : Virtues and Vices, Oxford, Basil Blackwell, 1978

Fotion, N.: "Paternalism" in Ethics, Vol. 89, 1979-1980, pp. 191-198

Gert, B. and Culver, C.M.: "The Justification of Paternalism", in Ethics, Vol. 89, 1979, pp. 199-210

Gert, B. and Culver, C.M.: "Paternalistic Behaviour", in Cohen, M., Nagel, T. and Scanlon T., 1981, pp. 201-213

Goldman, A.I. and J. Kim (eds.): Values and Morals, Dordrecht-Holland, D. Reidel Publishing Company, 1978

Gorman, J.L.: "Philosophical Confidence" in Evans, J.D.G. (ed.) 1987, pp. 71-79

Graham, G.: Contemporary Social Philosophy, Oxford, Basil Blackwell, 1988

Gray, J.: Mill on Liberty: A Defense, London, Routledge and Kegan Paul, 1983

Gustafson, D. (ed.): Essays in Philosophical Psychology, London, Macmillan, 1964

Griffiths, A.P. ed.: Of Liberty, Cambridge, Cambridge University Press, 1983

Haldane, J.: "Value Realism, Supervenience and Projectivism", photocopy, no year.

Hare, R.M.: The Language of Morals, Oxford, Clarendon Press, 1952

Hare, R.M.: Freedom and Reason, Oxford, Clarendon Press, 1963

Hare, R.M.: "Descriptivism" in Hudson (ed.): The Is-Ought Question, London, Macmillan, 1969, pp. 240-258

Hare, R.M.: "Universalizability" in Essays on the Moral Concepts, London, Macmillan, 1972, pp. 13-28

Hare, R.M.: Moral Thinking, Oxford, Clarendon Press, 1981

Hare, R.M.: "Ontology in Ethics" in Honderich (ed.), 1985, p. 39-53

Hare, R.M.: Essays in Ethical Theory, Oxford, Clarendon Press, 1989

Hare, R.M.: "Supervenience" in Hare, R.M. 1989

Harman, G.: "Human Flourishing, Ethics, and Liberty" in Philosophy and Public Affairs, Vol. 12, No. 4, Fall 1983, p.307-322

Hodson, J.D.: "The Principle of Paternalism", in American Philosophical Quarterly, Vol. 14, 1977, pp. 61-69

Holland, R.F.: Against Empiricism, Oxford, Basil Blackwell, 1980

Honderich, T.(ed.): Morality and Objectivity, London, Routledge & Kegan Paul, 1985

Hookway, C.: "Two Conceptions of Moral Realism" in The Aristotelian Society, Supplementary Volume, 1986, p. 189-205

Hudson, W.D.(ed.): The Is-Ought Question, London, Macmillan, 1969

Hurka, T.: ""Good" and "Good For"" in Mind, Vol. XCVI, No. 381, January 1987, pp. 71-73

Hurka, T.: "The Well-rounded Life" in The Journal of Philosophy, Vol. LXXXIV, No. 12, December 1987, pp. 727-746

Husak, D.N.: "Paternalism and Autonomy", in Philosophy and Public Affairs 1980, Vol. 10, No. 1, pp. 27-46

Kant, I.: Groundwork of the Metaphysic of Morals, in H.J. Paton, 1948

Karlsson, M.: "Epistemic Leaks and Epistemic Meltdowns" in Hume Studies, Vol. XIV, Nr. 2, November 1990, pp. 121-130

Laslett P. and Fishkin J. (eds.): Philosophy, Politics and Society, Oxford, Basil Blackwell, 1979

Letwin, O.: Ethics, Emotion and The Unity of The Self, Croom Helm, London, 1987

Lindley, R.: Autonomy, London, Macmillan, 1986

Lively, J.: "Paternalism", in Of Liberty, Griffiths, A.P. (ed.), Cambridge, Cambridge University Press, 1983, pp. 147-165

McConnell, T.C.: "Objectivity and Moral Expertise", in Canadian Journal of Philosophy, Vol. XIV, Nr. 2, June 1984, pp.193-216

MacIntyre, A.: After Virtue, London, Duckworth, 1985

Mackie, J.: Ethics, Harmondsworth, Penguin Books, 1977

McDowell, J.: "Are Moral Requirements Hypothetical Imperatives?", in The Aristotelian Society, Supplementary Volume 1978, pp. 13-29

McDowell, J.: "Non-Cognitivism and Rule-Following", in Holtzman and Leich (eds.): Wittgenstein: To Follow a Rule, London, Routledge & Kegan Paul, 1981, pp. 141-162

McDowell, J.: "Values and Secondary Qualities" in Honderich (ed.) 1985, pp. 110-129

McFetridge, I. G.: "Are Moral Requirements Hypothetical Imperatives?", in The Aristotelian Society, Supplementary Volume 1978, pp. 31-42

McGinn, C.: "Semantics for Nonindicatives" in Philosophical Studies, 1977, Nr. 32, pp. 301-311

McGinn, C.: The Subjective View, Oxford, Clarendon Press, 1983

McNaughton, D.: Moral Vision, Oxford, Basil Blackwell, 1988

Mill, J.S.: "On Liberty" in Three Essays, Oxford, Oxford University Press, 1975, pp.5-141

Mill, J.S.: "Utilitarianism" in Utilitarianism Warnock, M. (ed.) London, Fontana, 1962, pp. 251-321

Murdoch, I.: The Sovereignty of Good, Routledge and Kegan Paul, London, 1970

Nagel, T.: The Possibility of Altruism, Princeton, Princeton University Press, 1970

Nagel, T. : Mortal Questions, Cambridge, Cambridge University Press, 1979

Nagel, T. : The View from Nowhere, New York, Oxford University Press, 1986

Nielsen, K.: "On Being Morally Authoritative", in Mind, Vol. 89, 1980, pp. 423-430

Nowell-Smith, P.H.: Ethics, Penguin Books, Harmondsworth, 1954.

Nozick, R.: Anarchy, State and Utopia, Basil Blackwell, Oxford, 1974

Nozick, R.: Philosophical Explanations, The Belknap Press of Harvard University Press, Cambridge, Massachusetts, 1981.

O'Neill O.: "Abstraction, Idealization and Ideology in Ethics" in Evans J.D.G. (ed.), 1987, pp. 55-69

O'Neill, O.: "The Power of Example", in O'Neill, O., 1989

O'Neill, O.: Constructions of Reason, Cambridge, Cambridge University Press, 1989

Paskins, B.: "Philosophy in the Nuclear Age" in Evans J.D.G. (ed.), 1987, pp. 95-116

Paton, H.J.: The Categorical Imperative, London, Hutchinson, 1948

Paul, J. and Miller, F.D. jr.: "Communitarian and Liberal Theories of the Good", in Review of Metaphysics, Vol. 43, 1990, pp. 803-830

Pierce, C.: "Hart on Paternalism", in Analysis, Vol. 35, 1975, pp. 205-207

Platts, M. : "Moral Reality and The End of Desire", in M. Platts, 1980, pp. 69-82

Platts, M. (ed.) : Reference, Truth and Reality, London, Routledge and Kegan Paul, 1980

Putnam, H.: Meaning and the Moral Sciences, Routledge and Kegan Paul, London, 1978

Putnam, H.: Reason, Truth and History, Cambridge University Press, Cambridge, 1981

Putnam, R. A.: "Creating Facts and Values" in Philosophy, Vol. 60, 1985, pp. 187-204

Quine, W.V.: "On The Nature of Moral Values" in A.I. Goldman and J. Kim (eds.), 1978, pp. 37-45

Quinn, W.: "Reflection and the Loss of Moral Knowledge: Williams on Objectivity" in Philosophy and Public Affairs, Vol. 16, Nr. 2, 1987, pp.195-209

Rachels, J. ed.: Moral Problems, New York, Harper & Row, 1979

Railton, P.: "Alienation, Consequentialism, and The Demands of Morality" in Scheffler, S. (ed.) 1988. pp. 93-133; appeared originally in Philosophy and Public Affairs, Vol. 13, No. 2, 1984

Railton, P.: "Moral Realism" in Philosophical Review, Vol. XCV, 1986, 163-207

Rasmussen, S.A.: "Quasi-Realism and Mind-Dependence", in Philosophical Quarterly, Vol. 35, 1985, pp. 185-191

Rawls, J.: A Theory of Justice, Oxford, Oxford University Press, 1971

Regan, D.H.: "Freedom, Identity and Commitment" in Sartorius, R. (ed.), 1983, pp. 113-138

Reinhardt, L.: "Radical Freedom", in Philosophy, Vol. 60, 1985, pp. 89-104

Ryan, A. (ed.): The Idea of Freedom: essays in honour of Isaiah Berlin, Oxford, Oxford University Press, 1979

Ryle, G. : "Pleasure", in D. Gustafson (ed.), 1964, pp. 195-205

Sartorius, R. (ed.): Paternalism, Minneapolis, University of Minnesota Press, 1983

Sayre-McCord, G. (ed.): Essays on Moral Realism, Ithaca, Cornell University Press, 1988

Sayre-McCord, G.: "Introduction: The Many Moral Realisms" in Sayre-McCord, G. (1988), pp. 1-23

Scanlon, T.M.: "Rights, Goals and Fairness" in Scheffler, S. (ed.) 1988, pp. 74-92; appeared originally in Hampshire, S. (ed.): Public and Private Morality, Cambridge, Cambridge University Press, 1978

Scheffler, S. (ed.): Consequentialism and Its Critics, Oxford, Oxford University Press, 1988

Schueler, G.F.: "Modus Ponens and Moral Realism", in Ethics, April 1988, Vol 98, pp. 492-500

Scruton, R.: "Freedom and Custom", in Griffiths, A.P. (ed.), 1983, pp.181-196

Singer, P.: "Moral Experts", in Analysis, Vol. 32, 1972, pp. 115-117

Singer, P. and Wells, D.: The Reproductive Revolution, Oxford, Oxford University Press, 1984

Skorupski, J.: John Stuart Mill, Routledge, London, 1989

Skorupski, J.: "Liberal Elitism", a photocopy, no year.

Smart, J.J.C. and Williams, B.: Utilitarianism: for and against, Cambridge, Cambridge University Press, 1973

Solomon, W.D.: "Moral Realism and Moral Knowledge", in Proceedings of The American Catholic Philosophical Association, Vol. 59, pp. 41-57, 1985

Sturgeon, N.L.: "Harman on Moral Explanations of Natural Facts" in The Southern Journal of Philosphy, 1986, Vol. XXIV, Supplement, pp. 69-78

Sturgeon, N.L.: "What Difference Does It Make Whether Moral Realism Is True?", in The Southern Journal of Philosophy, 1986, Vol. XXIV, Supplement, pp. 115-141

Sturgeon, N. L.: "Moral Explanation" in Sayre-McCord, G. 1988, pp. 229-255

Swinburne, R.G.: "The Objectivity of Morality", in Philosophy, Vol. 51, 1976, pp. 5-20

Szabados, B.: "On 'Moral Expertise'", in Canadian Journal of Philosophy, Vol. VIII, Nr. 1, March 1978 pp. 117-129

Taylor, C.: "What's Wrong with Negative Liberty", in Ryan, A. (ed.) 1979, pp. 175-193

Ten, C.L.: Mill on Liberty, Oxford, Oxford University Press, 1980

Walker, R.C.S.: The Coherence Theory of Truth, London, Routledge, 1989

Walzer, M.: Spheres of Justice, New York, Basic Books, 1983.

Warnock, M.: A Question of Life, Oxford, Basil Blackwell, 1985

Werner, R.: "Ethical Realism", in Ethics, July 1983, Vol. 93, pp. 653-679

White, M.: What I and Ought to Be Done, New York, Oxford University Press, 1981

Wiggins, D.: "Truth, Invention and the Meaning of Life" Proceedings of the British Academy, Vol. LXII, 1976; also in Wiggins, 1987, pp. 87-137

Wiggins, D.: Needs, Values, Truth: Essays in the Philosophy of Value, Oxford, Basil Blackwell, 1987

Wikler, D.: "Paternalism and The Mildly Retarded" in Sartorius, R. (ed), 1983, pp. 83-94

Williams, B.: Morality, Cambridge University Press, Cambridge, 1972

Williams, B.: Moral Luck, Cambridge, Cambridge University Press, 1981

Williams, B.: "Ethics and the Fabric of the World" in Honderich (ed.), 1985, pp. 203-214

Williams, B.: Ethics and the Limits of Philosophy, London, Fontana Press, 1985

Winch, P.: Ethics and Action, London, Routledge and Kegan Paul, 1972

Woozley, A.D.: Euthanasia and the Principle of Harm" in Rachels, J. (ed.), 1979, pp. 498-507

Wright, C.: "Realism: The Contemporary Debate - Whither Now?" photocopy of a contribution to the conference Realism and Reason, St. Andrews, 1988

Wright, C.: "Moral Values, Projection and Secondary Qualities", in The Aristotelian Society, Supplementary Volume 1988, p. 1-26

von Wright, G.H.: The Varieties of Goodness, London, Routledge and Kegan Paul, 1963